1992

Cognitive Evaluations Using Human Figure Drawings: An Empirical Investigation of Two Methods

Steven C. Abell
Loyola University Chicago

Cognitive Evaluations Using Human Figure Drawings: An Empirical Investigation of Two Methods

by

Steven Abell

Dissertation Submitted to the Faculty of the Graduate School of Loyola University of Chicago in Partial Fulfillment of the Requirements of the Degree of

Doctor of Philosophy

January

1992

Steven C. Abell

Loyola University Chicago

COGNITIVE EVALUATIONS USING HUMAN FIGURE DRAWINGS: AN EMPIRICAL INVESTIGATION OF TWO METHODS

Psychologists have often used human figure drawings as a measure of intelligence. Scoring systems have been developed by Buck (1948) and Goodenough and Harris (1963), to assess the level of cognitive ability that is demonstrated in human figure drawings. It was uncertain, however, which of these systems would most accurately assess the intellectual ability of adults. It was also unclear if a particular factor of intelligence, such as field independence, would be more related to human figure drawing performance than would overall IQ scores.

In order to address these questions, this investigation compared the performance of 101 normal adults on the Draw-A-Person test and the Wechsler Adult Intelligence Scale-Revised (WAIS-R). The first figures of the Draw-A-Man protocols were scored with the person component of Buck's (1948) House-Tree-Person Technique, as well as with the Goodenough-Harris Drawing Test.

It was expected that both Draw-A-Person scores would be

significantly related to Performance, Verbal, and Full Scale IQ scores on the WAIS-R. It was also expected that there would not be a significant difference in the level of relationship between the two Draw-A-Man scoring systems and Performance IQ, Verbal IQ, and Full Scale IQ. It was predicted as well that the factor of field independence, as measured by the Picture Completion, Block Design, and Object Assembly subtests of the WAIS-R, would be significantly more correlated with drawing performance that would overall IQ scores.

Results indicated that scores on both Buck's (1948) technique and the Goodenough-Harris Drawing Test were significantly related to Performance IQ and Full Scale IQ, but not to Verbal IQ. As expected, there was no significant difference in the level of relationship between the two Draw-A-Person scoring systems and Performance IQ, Verbal IQ, or Full Scale IQ. Results failed to validate the study's prediction that field independence would be more related to Draw-A-Person performance than would overall IQ scores.

Department: Psychology

Committee: Dr. James Johnson (Director),

Dr. Alan DeWolfe, Dr. Richard Maier

ACKNOWLEDGEMENTS

The author wishes to acknowledge the tremendous amount of assistance that he received from the director of this dissertation, Dr. James Johnson. He provided invaluable help with both the conceptualization and the actual completion of this project. Dr. Johnson's time and energy were greatly appreciated.

The author would also like to thank the other two members of his committee, Dr. Alan DeWolfe and Dr. Richard Maier. Their comments and advice were very valuable to the author.

Several other individuals deserve special recognition as well. Anna Heiberger gave the author a tremendous amount of practical assistance with data collection and analysis, as well as offering a great deal of encouragement and moral support during the entire project. Dr. Grayson Holmbeck generously made the archives of the Loyola University Test Library available to the author. Finally, Catharine Barnett provided invaluable assistance with the scoring and coding of data, and her contribution to the study is greatly appreciated as well.

VITA

The author, Steven Abell, was born on April 26, 1962, in Oklahoma City, Oklahoma.

Mr. Abell received his undergraduate education at Columbia University in the City of New York. He graduated Summa cum Laude from Columbia University in May of 1984, with the degree of Bachelor of Arts in English.

In September of 1985, Mr. Abell began graduate work in clinical psychology, at Loyola University Chicago. He received the degree of Master of Arts from Loyola University in May of 1988. Mr. Abell completed his internship in clinical psychology at Northwestern University Medical School.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS ii

VITA iii

LIST OF TABLES vi

Chapter

I. INTRODUCTION 1

Rationale for the Study. 1
Specific Aim 2
Overview of the Study. 3

II. REVIEW OF THE RELEVANT RESEARCH LITERATURE 6

The Clinical Beginnings. 6
The Preponderance of Empirical Research. . . . 11
Reliability 11
Administration 19
The Body Image Hypothesis. 22
Research of Self Esteem. 25
Sex and Gender Roles 29
Sexual Attitudes, Feelings, and Dysfunction. . 32
Other Psychological Variables. 36
Cultural and Environmental Factors 42
Structural and Formal Aspects. 46
Artistic Ability: A Possible Confound. 63
D-A-P Performance and Learning Disabilities. . 65
The Scoring of Cognitive Ability 68
Hypotheses 75

III. METHOD 78

Subjects 78
Measures 79
Procedure. 79

IV. RESULTS. 82

Page

Hypothesis #1. 83
Hypothesis #2. 87
Hypothesis #3. 92
Hypothesis #4. 97
Hypothesis #5. 101

V. DISCUSSION 109

Inter-Rater Reliability. 109
Validity of Buck's Scoring System. 110
Validity of the Goodenough-Harris Test 115
Buck vs. Goodenough-Harris: A Comparison . . . 118
Suggestions for Future Research. 119

REFERENCES 122

LIST OF TABLES

Table — Page

1. Mean Scores and Standard Deviations for Human Figure Drawing Scores and WAIS-R Scores. 84

2. Pearson Correlation Coefficients for All Subjects for Buck's (1948) Scoring System with Wechsler IQ Scores. 86

3. Direct Difference t-tests for Correlated Observations for Buck's (1948) Standard Scores with IQ Scores. 88

4. Pearson Correlation Coefficients for All Sugjects for Goodenough-Harris Scores with Wechsler IQ Scores 90

5. Direct Difference *t*-tests for Correlated Observations for Goodenough-Harris Standard Scores with IQ Scores. 91

6. Estimates of Omega Squared (ω^2) for Standard Scores with Performance and Full Scale IQ Scores 93

7. The *t*-test Values Obtained for the Difference in the Pearson Correlation Coefficients of the Goodenough-Harris Drawing Test with Performance IQ and Buck's (1948) Scoring System with Performance IQ. 94

8. The *t*-test Values Obtained for the Difference in the Pearson Correlation Coefficients of the Goodenough-Harris Drawing Test with Verbal IQ and Buck's (1948) Scoring System with Verbal IQ 95

Table Page

9. The t-test Values Obtained for the Difference in the Pearson Correlation Coefficients of the Goodenough-Harris Drawing Test with Full Scale IQ and Buck's (1948) Scoring System with Full Scale IQ 96

10. The t-test Values Obtained for the Difference in the Pearson Correlation Coefficients of Buck's (1948) Scoring System with Performance, Verbal, and Full Scale IQ Scores. . . 98

11. The t-test Values Obtained for the Difference in the Pearson Correlation Coefficients of the Goodenough-Harris Drawing Test with Performance, Verbal, and Full Scale IQ Scores. . . 100

12. Results of Regression Procedure with Goodenough-Harris Raw Score as Dependent Variable 104

13. Results of Regression Procedure with Goodenough-Harris Standard Score as Dependent Variable. 105

14. Results of Regression Procedure with Buck's (1948) Raw Score as Dependent Variable 106

15. Results of Regression Procedure with Buck's (1948) Weighted Score as Dependent Variable 107

16. Results of Regression Procedure with Buck's (1948) Standard Score as Dependent Variable. 108

CHAPTER I

INTRODUCTION

Rationale for the Study

Psychologists are often asked to make rapid assessments of an individual's cognitive abilties. While individual IQ tests, such as the Wechsler Adult Intelligence Scale-Revised, are generally regarded as the preferred method of assessing cognitive ability, the administration of these tests is not always possible. Individual IQ tests are relatively expensive and time-consuming to administer, and some individuals may be either unwilling or unable to take these measures. Human figure drawings, on the other hand, are quick and simple to administer. For this reason, many psychologists have found it useful to score human figure drawings for cognitive ability.

Scoring systems have been developed by Buck (1948) and Goodenough and Harris (1963), which attempt to assess the level of cognitive ability that is demonstrated in human figure drawings. Before the beginning of this study, however, a number of questions remained unanswered about the validity of these systems. For instance, it was not clear which of the systems could most accurately assess the

cognitive ability of adults. Past investigators had not compared the two systems, because Goodenough and Harris developed their system with children, while Buck developed his system with adults. But the items on the two scoring systems are quite similar, and may be able to reflect adult ability in the same manner.

It was also unclear what aspects of intellectual ability were related to these scoring systems. While both systems have been significantly related to overall IQ scores on tests such as the Wechsler-Bellevue, it was not known if certain IQ subtest scores would be more related to human figure drawing scores than would other subtest scores.

Specific Aim

This investigation sought to further our understanding of the value that human figure drawings may have in the assessment of cognitive ability. In particular, this study compared the performance of a normal adult population on the Draw-A-Person test and the Wechsler Adult Intelligence Scale-Revised (WAIS-R). The first figures of the Draw-A-Person protocols were scored with the person component of Buck (1948)'s House-Tree-Person Technique, and the Draw-A-Man scoring system that was developed by Goodenough and Harris (1963).

It was expected that both of these Draw-A-Person scores would be significantly related to Full Scale IQ scores on the WAIS-R. It was also expected that there would not be a

significant difference in the level of relationship that was found between the two Draw-A-Person scoring systems and Full Scale IQ scores on the WAIS-R. This study also sought to determine what factors of intelligence were most correlated with performance on the Draw-A-Person scoring systems. It was expected that the factor of field independence, as developed by Witkin (1965) and measured by the Picture Completion, Block Design, and Object Assembly subtests of the WAIS-R, would be signficantly more correlated with drawing performance than would Verbal, Performance, or Full Scale IQ.

Overview of the Study

The test protocols of 100 subjects were selected at random, from the archives of the Loyola University Test Library. The subjects were undergraduates at Loyola University of Chicago who volunteered to take a battery of psychological tests in order to assist doctoral students with their training in clinical psychology.

Once these protocols were obtained, the investigator scored the first human figure drawing of each subject according to the Goodenough-Harris Drawing Test, and for the person component of Buck (1948)'s House-Tree-Person Test. The Verbal IQ, Performance IQ, Full Scale IQ, and 11 subtest scores from the Wechsler Adult Intelligence Scale-Revised were also recorded for each subject. A field independence score was determined for each subject, by adding the Object

Assembly, Block Design, and Picture Completion subtest scores of the WAIS-R.

After these scores were recorded, the Goodenough-Harris Drawing Test Scores and the House-Tree-Person scores were both correlated with Verbal IQ, Performance IQ, Full Scale IQ, and field dependence scores, to develop a better understanding of the relationship between human figure drawings and individual IQ test scores. A regression analysis was also performed to determine if any particular WAIS-R subtest or group of subtests could serve as a superior predictor of performance on the two systems of scoring cognitive development from human figure drawings. By performing these statistical procedures, it was hoped that the present study would provide answers to several pertinent questions about the value of human figure drawings in the assessment of cognitive skills.

Such questions about the use of human figure drawings in the assessment of cognitive skills, however, did not arise from a vacuum of knowledge. There has been a long tradition of clinical research about the usefulness of human figure drawings. To place the present study in its proper intellectual context, the empirical literature on the Draw-A-Person test will be reviewed. This review will include a discussion of the clinical origins of the Draw-A-Person Test, a review of the empirical research on its reliability and validity, the use of human figure drawings as a

projective measure, and a detailed discussion of the existing research on the use of human figure drawings as a measure of cognitive abilities.

CHAPTER II

REVIEW OF THE RELEVANT RESEARCH LITERATURE AND STATEMENT OF PROBLEM

The Clinical Beginnings

The idea, that spontaneous drawings could shed light on the character or abilities of the individual, is not a new one. In the late nineteenth century, educators were beginning to suspect that the drawings of children were often accurate reflections of their developmental level (Harris, 1963). While the field of psychology was in its infancy, Ebenezer Cooke (1885) published an article suggesting that children's drawings went through a series of successive changes as the children matured. Cooke, an educator, described the changes that he had observed in his classroom, and recommended that art education be designed to correspond to these changes. A few years latter, Ricci (1887), apparently unaware of Cooke's work, published a similar article on the developmental sequence of drawings by Italian school children.

As the young field of psychology developed, psychological researchers were quick to explore the developmental aspects of children's spontaneous drawings. A number of descriptive studies were conducted. Researchers

viewed drawings produced by children in each grade of school, and tried to form intuitive conclusions about how drawings changed as children got older. The first summary of this research was published by Cyril Burt (1921), in which it was suggested that children's drawings become increasingly "less primitive and savage like (Burt, p.69)" as children got older. For the most part, however, little was known about what particular aspects of human development children's drawings were supposed to represent (Harris, 1963).

It was Florence Goodenough (1926) who advanced the study of children's drawings by demonstrating how human figure drawings reflected intellectual development. Unlike earlier research on spontaneous drawings, Goodenough's work was psychometrically based, and has been successfully replicated in Europe, Africa, and Japan (Harris, 1963). Goodenough's chief contribution was to develop a reasonably reliable scoring system, based an the inclusion of realistic details about the human figure. Goodenough found that as children became more intellectually mature, they drew figures which were increasingly filled with realistic details such as shoelaces, shirt collars, and eyebrows.

With the development of Goodenough's system, clinicians began to use drawings as part of their standard test battery, and the Draw-A-Person test was born. Goodenough's system was particularly useful to clinicians in

underdeveloped parts of the world, "where convenient nonverbal measures were needed to classify large numbers of nonreading children for educational purposes" (Harris, 1963, p.11). While Goodenough concentrated on the intellectual component of children's drawings, she also recognized that the drawings revealed emotional maturity and psychopathology (Taylor, 1977).

As the use of human figure drawings became more widespread, Goodenough (1926) was not the only one to recognize the ways in which these drawings reflected the individual's emotional makeup. During the 1930's and 40's, the Draw-A-Person became increasingly popular as a projective measure of personality (Harris, 1963). In keeping with this new interest in the projective use of the Draw-A-Person, a number of manuals were published to guide the neophyte clinician in his or her attempts to understand the relationship between personality and figure drawings. Of the various interpretive manuals, John Buck (1948), Karen Machover (1949) and Emmanual Hammer (1954, 1958) are generally accepted as having developed the most influential systems.

It was Machover (1949) who particularly influenced the field, with her development of what has been termed "the body image hypothesis" (Swensen, 1968). Machover's basic premise was that when a person draws a human figure, the person also draws a picture of how he or she views him or

herself. In other words, without a conscious awareness of the process, individuals were thought to project the core elements of their personality into their drawings. Machover also believed that emotionally disturbed individuals, who lacked awareness of reality, would demonstrate their poor reality contact by including bizarre or inappropriate details in their drawings. In a related fashion, Machover believed as well that human figure drawings would reveal the individual's preferred defense mechanisms, be they projection, denial, or any other combination of psychological defenses. In general, Machover took an extremely optimistic stance about the ability of human figure drawings to reveal the individual's self-concept and emotional difficulties.

Hammer (1958) tended to share Machover's optimism. In the opening of his ground breaking book on the interpretation of projective drawings, Hammer proclaimed "show me what you draw and I'll tell you what you are" (p.5). Hammer goes on to suggest many of the numerous ways in which drawings can be used, and recommends that clinicians give an entire battery of drawings to their testing patients. In addition to the traditional figure drawings, Hammer suggests the House-Tree-Person developed by Buck (1948), the Draw-An-Animal, the Draw-A-Family, and the Draw-A-Person-in-the-Rain. Like Machover (1949), Hammer has an elaborate system for the interpretation of human figure

drawings, based on the notion that these drawings are disguised depictions of the individual's self-concept.

With the publication of Machover's (1949) and Hammer's (1958) manuals on the interpretation of projective drawings, these tests became exceedingly popular with clinicians. Sundberg (1961) discovered that, of the 12 tests most frequently used by the 185 clinical service agencies in the United States, 3 were projective drawing tests. Sundberg also reported that the Draw-A-Person was second only to the Rorschach in popularity. A few years latter, Wildman and Wildman (1967) surveyed 100 members of the American Psychological Association, and discovered that the Draw-A-Person had dropped to eighth place in overall test use, but was still fifth in popularity for quick evaluations. By the time Lubin, Wallis, and Paine (1971) surveyed 251 clinical service agencies in the United States, they found that the Draw-A-Person was still in the ten most frequently used tests, and had risen to fifth in overall popularity. While empirical data on the current popularity of the Draw-A-Person test is not available, Kahill (1982) conducted an informal telephone survey of clinical psychologists across North America. Kahill's results suggest that the Draw-A-Person remains one of the most frequently used psychological tests in this country. Given the popularity of the test, it is not surprising that a tremendous amount of empirical

research has attempted to investigate its reliability and validity.

The Preponderance of Research on Human Figure Drawings

An examination of the available research on the Draw-A-Person Test reveals that hundreds if not thousands of studies have been published on this topic. Both the reliability and validity of the Draw-A-Person Test have been examined in numerous ways by a myriad of researchers. At this point in time, the complete body of empirical literature on human figure drawings is perhaps too large and unwieldy to cover adequately in a single review.

Fortunately, several excellent reviews of the available research literature have been published. Swensen (1957, 1968) and Roback (1968) both reviewed the empirical investigations of Machover (1949)'s hypotheses that took place between the publication of Machover's book and the mid 1960's. Kahill (1984) then updated the work of Swensen and Roback by reviewing the available literature from 1968-1982. In order to build upon the work of these scholars, the present review will focus on summarizing the work of Swensen, Roback, and Kahill, and discussing research that has been published since 1982.

Reliability

Perhaps the logical place to begin a review of the available research of the Draw-A-Person Test, or of the research on any assessment device, would be with an overview

of the research on the measure's reliability. In classical psychometric thinking, the reliability of an instrument needs to be established before the measure can be of any potential use to clinicians (Anastasi, 1982).

Machover (1949) believed that structural and formal aspects of the Draw-A-Person, such as placement and shading, would be drawn consistently, while the content of the Draw-A-Person, such as clothing or facial expression, would be much less reliable. Machover made this prediction, because she believed that content was a reflection of the individual's current emotional condition. She believed that a person's emotions are likely to be relatively unstable, while the structural aspects of the drawings represent a person's cognitive skills, which are likely to be more consistent. Researchers who have attempted to test Machover's findings have been faced with two hurdles: developing a reliable scoring system which would provide adequate interrater reliability, and exploring Machover's ideas which pertained to the test-retest reliability of the drawings.

In general, early research findings were not consistent with Machover's (1949) predictions. After reviewing 16 studies, Swensen (1968) concluded that the reliability of any particular Draw-A-Person sign was a function of how much drawing behavior was included in that sign. For instance, global assessments of the overall quality of the drawings,

reliability in a study of 20 college students. Guinan and Hurley obtained drawings from these subjects on two separate occasions five weeks apart. They asked judges to match the drawings obtained on one occasion with those received on another occasion. There were three groups of judges: a group of Ph.D.'s, a group of graduate students, and a group of college freshmen. Results indicated that the judges were able to match the drawings significantly at the .001 level, and that the Ph.D.'s were correct on an average of better than 19 out of 20 matches (the freshmen were only correct on an average of 12 out of 20 matches). This study suggests that with explicit instruction or some training, judges can match figure drawings with satisfactory reliability.

It seems likely, however, that when judges match drawings, they attempt to use as much information from the drawings as possible. Swensen (1968) concluded that when clinicians attempt to make more discreet assessments of human figure drawings, both interrater and test-retest reliability decline. After reviewing many studies, Swensen found that global measures were the most reliable, with structural and formal aspects being somewhat less reliable, and content being the least reliable of all. Swensen explained this finding by suggesting that "the reliability of a particular sign is a direct, linear function of the amount of drawing behavior included to assess that sign"

(p.40). In other words, if the clinician asks a broad question about the sophistication of the drawings, he or she may obtain fairly reliable results. If the clinician, on the other hand, asks precise questions about content, such as does the figure wear a hat or carry a cane, he or she will have little chance of achieving adequate test-retest reliability. Since he believed strongly in this pattern of reliability correlations, Swensen concluded "that global ratings are the most reliable, and therefore the most useful aspect of the DAP" (p.40).

In a separate review of empirical research on the Draw-A-Person, Roback (1968) agreed with Swensen's conclusion that global ratings of human figure drawings were the most reliable. Roback, however, offered readers a different explanation for this finding. Roback suggested that the poor performance of structural and content signs may have been due to the methods of researchers, rather than to a linear relationship between the amount of drawing behavior used to assess a sign and the reliability of that sign. Roback concluded that while empirical research had "generally failed to support Machover's hypotheses (1949), there is still an insufficient number of well-designed investigations from whose findings it could be concluded that 'the patient (the DAP) died' " (p.16). Roback made several specific suggestions about how to improve Draw-A-Person research in the future. Roback believed that the

(p.40). In other words, if the clinician asks a broad question about the sophistication of the drawings, he or she may obtain fairly reliable results. If the clinician, on the other hand, asks precise questions about content, such as does the figure wear a hat or carry a cane, he or she will have little chance of achieving adequate test-retest reliability. Since he believed strongly in this pattern of reliability correlations, Swensen concluded "that global ratings are the most reliable, and therefore the most useful aspect of the DAP" (p.40).

In a separate review of empirical research on the Draw-A-Person, Roback (1968) agreed with Swensen's conclusion that global ratings of human figure drawings were the most reliable. Roback, however, offered readers a different explanation for this finding. Roback suggested that the poor performance of structural and content signs may have been due to the methods of researchers, rather than to a linear relationship between the amount of drawing behavior used to assess a sign and the reliability of that sign. Roback concluded that while empirical research had "generally failed to support Machover's hypotheses (1949), there is still an insufficient number of well-designed investigations from whose findings it could be concluded that 'the patient (the DAP) died' " (p.16). Roback made several specific suggestions about how to improve Draw-A-Person research in the future. Roback believed that the

"necessity for more objective scoring procedures" (p.16) was shown by the poor interrater reliability correlations obtained by most research teams for structure and content. For structural variables, Roback found coefficients that were mostly .30 to .50, and for content, coefficients were generally unavailable and were expected to be even lower. Roback hypothesized that these dismal results were due simply to the sloppy, idiosyncratic manner in which human figure drawings were scored. Roback was also convinced that researchers made little attempt to assess the emotional condition of their subjects in a psychometrically sound fashion. He indicated that "clinical criteria such as neurosis, psychosis, conduct disturbance, and maladjustment have been accepted as ready made psychiatric labels without any further refinement" (Roback, p.17). When these different problems occured simultaneously, Roback believed that researchers ended up trying to relate poorly scored figure drawings to relatively meaningless psychiatric labels.

Kahill (1984)'s more recent review of Draw-A-Person research suggests that investigators did heed some of Roback (1968)'s criticisms. Kahill reviewed nine studies which were published between 1968 and 1982, and found:

> the interrater reliabilities of both content and structural/formal elements to be equal to or better than those of global ones. Perhaps the realization of the relatively poor performance of these more limited aspects of figure drawings has led to an increased motivation to objectify and

> standardize rating procedures and to adequately train judges, with a corresponding increase in reliability. (p.271)

The work of Maloney and Glasser (1982) would seem to support Kahill's idea that an increased emphasis upon the training of judges can greatly enhance reliability. By comparing a group of judges who were pre-trained with a comparable group of judges who received no pre-training, Maloney and Glasser demonstrated that pre-training can significantly increase interrater reliability. By taking the time to adequately pre-train judges on more objective scoring systems, researchers appear to have dramatically improved interrater reliability between 1968 and 1982.

Only four studies attempted to examine test-retest reliability during those years. In general, these studies found adequate levels of test-retest reliability. A test-retest $\underline{r}$ of .81 was reported for proportional accuracy (Beck & Bart, 1970), .88 for a global body image scale (Maloney & Payne, 1969), and .90-.99 for overall quality (Jensen, Prandoni, & Abudabbeh, 1971). Marzolf and Kirchner (1970) reported their test-retest data for six content variables in terms of the percentage of their subjects who drew the same signs on two separate occasions. Their findings were 63.5% consistency for ears, 66.7% for hands, 68.9-75.7% for facial expression, 77.5% for feet, 83.6% for person same sex, and 83.9% for eyes. While the results of the four research teams just described cannot be considered definitive, they

did provide the field with further encouraging evidence about the potential reliability of the Draw-A-Person test.

Shortly after Kahill (1984) concluded her review of the literature in 1982, Rubin, Schachter, and Ragins (1983) provided the field with an interesting new twist. Rubin et al. obtained human figure drawings from 180 school children on two consecutive days, and then the following week on two more consecutive days. Rubin et al. compared figure drawings on consecutive days, as well as those drawn a week apart. The drawings were scored for intelligence using the Goodenough-Harris rating system, for visual similarity, and for content. Rubin et al. found "significantly less intra-individual variability in the area of content than in either intelligence scores or visual similarity scores" (p.660). In this case, it was actually content which appeared to be the most reliable variable across the four different test administrations. Rubin et al. commented on how this finding is in contrast to the hypothesis of Machover (1949), that "structural and formal aspects of the drawing are less subject to variability than content" (Machover, p.17). Rubin et al. explained their finding as a product of the global manner in which they chose to score content. The content of each drawing was scored as either man, woman, boy, or girl. Finer distinctions such as "Grandfather and Indian" (Rubin et al., p.660) were ignored. This again suggests, that the manner in which the scoring system is

devised will play a large role in determining the reliability of the Draw-A-Person.

Administration

In a similar manner, it seems likely that the way in which the test is administered will also influence its results. Unfortunately, empirical researchers have paid scant attention to the administration of the Draw-A-Person in the past. Neither Roback (1968) nor Kahill (1984) mention this factor in their exhaustive reviews of the available research literature. Since the Draw-A-Person is a relatively simple measure to administer, researchers may have erroneously believed that everyone would administer the test in a standard fashion. A cursory glance of several different Draw-A-Person manuals, however, reveals that their instructions are not always similar. Machover (1949) instructs subjects to "draw a person. Draw the best person you can. Make your drawing a whole person and not a stick figure" (p.32). Goodenough and Harris (1963), on the other hand, instruct children to draw an entire body and to draw one gender or the other. Their typical instructions read: "Draw a picture of a man. Make the very best picture you can. Be sure to make the whole man, not just his head and shoulders" (Goodenough & Harris, p.1). Goodenough and Harris found that young children tended to draw only faces, so they altered their instructions to change this situation.

Ponzo (1957) also altered the standard instructions, by asking subjects to draw figures "like an idiot would" (p.278). Ponzo found that the "idiot" drawings were rated by observers as more primitive and careless. Ponzo interpreted this to mean that his manipulation of the instructions made subjects feel less emotionally inhibited.

In a similar way, Pfeffer (1987) found that slight changes in the standard instructions could alter the ethnic identity of children's drawings. Pfeffer (1987) based her study on the research of Schofield (1978) and on an earlier investigation which she conducted herself (Pfeffer, 1984). Schofield found that when African-American children were told to draw a person, they predominately drew caucasian figures, and Pfeffer (1984) obtained similar results with Yoruba children in Nigeria. But when Pfeffer (1987) later told a group of 134 Yoruba children to "draw yourself" (p.780), she found that these children drew significantly darker figures than did children told simply to "draw a person."

Even when the instructions are standardized, other aspects of the test's administration can still influence the final outcome. Cassel, Johnson, and Burns (1958) studied the performance of subjects with the examiner either present or absent, and discovered that more deviant signs (according to Machover's manual) were present when the examiner left the room. West, Baugh, and Baugh (1963) found that under

hypnosis, subjects drew smaller and more developmentally immature drawings.

Many researchers have also wondered whether or not the sex of the administrator would influence the sex of the first drawn figure. The answer appears to be a somewhat qualified no. Research indicates that the administrator's gender has little effect on the performance of children (Datta & Drake, 1968) or adults (Holtzman, 1952), when the test is given individually. Bauer and Paludi (1979) on the other hand, found that the administrator's sex was a factor when the test was given to undergraduates in a group format. Using same sex groups, Brauer and Paludi tested groups of men and groups of women using administrators of the same and of the opposite sex. They found a positive correlation between the sex of the administrator and the first-drawn sex; for example, a group of men tested by a woman drew 47% more opposite sex figures than a group of men tested by a man. Jensen (1985), however, tested groups of male and female undergraduates with both male and female administrators, and failed to replicate Brauer and Paludi's results. Jensen reports that the sex of the administrator failed to produce a significant difference in the sex of the first drawn figure. Jensen speculates that in the somewhat unnatural environment of a single sex group, the sex of the administrator may become a particularly salient fact to subjects, with the administrator then serving as an

influential role model. Otherwise, the sex of the administrator seems to make little difference in the drawings that are produced.

Levy and Barowsky (1986) found that the mode in which the Draw-A-Person test is administered may make little difference as well. Levy and Barowsky compared computer-administered Harris-Goodenough Draw-A-Man test protocols with protocols obtained from standard paper and pencil administrations. For the computer assisted administration, subjects were given "an Apple IIe computer, equipped with a pressure-sensitive template commercially available as the Koala Pad and stylus, as well as the necessary pre-programmed software" (Levy & Barowsky, p.396). The only difference obtained, according to the Goodenough-Harris scoring system, was that female subjects tended to produce significantly more body and clothing details on the paper and pencil administration. Otherwise, the results obtained by Levy and Barowsky suggest that performance on the Draw-A-Person is relatively stable across different modes of administration.

The Body Image Hypothesis

In addition to issues concerning the administration and reliability of the Draw-A-Person, the validity of this test is also an issue of importance. In the past, the central idea for researchers to examine has been the body image hypothesis. It was Machover (1949) who first stated the

body image premise, when she indicated that:

> the human figure drawn by an individual who is directed to "draw a person" relates intimately to the impulses, anxieties, conflicts, and compensations characteristic of that individual. In some sense, the figure drawn is the person, and the paper corresponds to the environment. (p.35)

Over the years, this idea has become known as the body image hypothesis, and is considered the theoretical underpinning for the Draw-A-Person test. Hammer (1958) supported the body image hypothesis by quoting Elbert Hubbard, who stated that "when an artist paints a portrait, he paints two, himself and the sitter" (Hammer, p.8).

As researchers have tested the body image hypothesis, they have actually attempted to test two separate notions. First, researchers have explored whether or not subjects project their physical identity into their drawings. In other words, do subjects draw figures who match their weight, height, race, and facial features? As a second area of inquiry, researchers have also attempted to explore whether or not subjects project their personality and emotions into the drawings. Do aggressive individuals draw aggressive figures? Do depressed individuals draw a certain type of portrait? These questions have occupied psychological researchers for the past four decades.

In their extensive reviews of the published research literature, Swensen (1968), Roback (1968), and Kahill (1984) all report mixed findings in regards to the body image

hypothesis. Swensen reviewed numerous studies conducted between the years 1957-1966, and found:

> whether on not the Draw-A-Person reflects a subject's concept of his own body will be difficult to determine, and perhaps not necessarily meaningful. But the data does indicate that scores on various aspects of the DAP are significantly related to some other measures that would be expected to reflect a subject's image of himself. (p.24)

Swensen also comments on the fact that most studies using adult subjects found some significant relationship between the Draw-A-Person and some other measure of body image or self-concept. Studies with children, on the other hand, generally failed to discover such relationships. This led Swensen to conclude that Draw-A-Person protocols may mean one thing for adults, and an entirely different thing for children.

Like Swensen (1968), Roback (1968) also reviewed the studies on the Draw-A-Person that were published approximately up to 1967. While Roback limited his review of the literature to studies of adults, he did examine research which covered the relationship of both body image and self concept to Draw-A-Person performance. Roback ultimately echoed Swensen's conclusion, when Roback stated that "although there appears to be some support for Machover's hypotheses, the inconsistent findings indicate that the relationship between figure drawings and body image is still unclear" (p.3).

When Kahill (1984) attempted to update the work of Swensen (1968) and Roback (1968), Kahill reported that the results "were quite mixed, with slightly more findings failing to support the body image hypothesis than supporting it" (p.271). Kahill concluded that there is probably some relationship between self-concept and human figure drawings. The problem is that we often tend to interpret the notion of self in too restricted of a fashion. Kahill suggests that "the projection of self should not be narrowly defined, and that it may involve one's actual self, one's idealized self, one's feared self, and one's perception of significant others" (p.274). Kahill believes that the existing body of research suggests that Hammer (1959) was correct, when Hammer suggested that the Draw-A-Person taps an extremely wide array of feelings about oneself and others. The task for researchers is both to determine if the Draw-A-Person is related to self-concept, and to determine what specific aspects of self-concept may be projected into human figures.

Research on Self-Esteem

Since Kahill (1984) published her review, researchers have continued to focus on the central aspects of identity that may be projected into human figure drawings. Numerous areas of psychological identity have been explored, with self-worth or self-esteem being one of the most prevalent.

Paine, Alves, and Tubino (1985) found that pediatric oncology patients drew self drawings significantly smaller

in size than those of their healthy peers, suggesting that the smaller size drawings reflected lower self-esteem associated with their physiological disease. In a somewhat similar fashion, Hamilton (1984) suggests that human figure drawings can reflect the self-concept development of children who participate in bilingual education. For three years, Hamilton followed a group of children who participated in a bilingual education program in El Paso, Texas. The children, who ranged from five to eight in age, were all Spanish language dominant and from economically disadvantaged families. Hamilton used repeated administrations of the human figure drawings test as his measure of self-esteem, and reports that these children experienced significant gains in self-concept during the three year period. This appeared to demonstrate the worth of bilingual education, since children from impoverished, Spanish language dominant families generally suffer from decreases in their self-concept during the first few years of school. Hamilton's results are questionable, however, given the loose way in which he measured self-esteem. Hamilton developed rather vague guidelines, in which larger drawings, with more elaborate details, were considered indicative of high self-esteem. Three judges then evaluated the drawings on a scale of 1 to 9, using Hamilton's criteria. No attempt was made to assess how judges weighed the different aspects of the drawings, such as size or level

of details, in their decisions about overall self-concept. It was simply assumed, that larger and more elaborate drawings would be clearly indicative of positive self-esteem.

The work of Calhoun, Ross, and Bolton (1988) suggests that this may not be the case. Calhoun et al. found that self-esteem was negatively related to the performance of 9 to 12 year old boys on the Goodenough-Harris Draw-A-Man Test (which awards more points to drawings with more details). Calhoun et al. gave the Coopersmith Self-Esteem Inventory to fourteen boys, and found that those with lower self-esteem scores included more details in their drawings, with a highly significant correlation of .71. Interestingly, these researchers found no relationship between self-esteem and the Goodenough-Harris Test for girls. They also found that the girls in their sample had significantly higher self-esteem scores. Calhoun et al. note that all of the children in their study had female teachers, and two-thirds were from households headed by women. They hypothesize that the lack of male role models may have lead the boys to suffer from lower self-esteem. Calhoun et al. report that many of the highly elaborate portraits done by boys were of male fantasy figures, such as "cowboys, karate fighters, cartoon heroes, etc." (p.254), suggesting that the boys developed these detailed drawings to compensate for their poor self-esteem.

It may be that boys demonstrate positive self-esteem not through the use of elaborate details, but simply through the size of their figures. Delatte and Hendrikson (1982) found that size of human figure drawings and self-esteem were significantly correlated for adolescent boys, but not for girls. On the other hand, Delatte (1985) later found that positive self-esteem for adolescent girls was related to the femininity of their female drawings, rather than to the size of the figures. Delatte (1985) obtained a significant correlation coefficient of .31, for the femininity ratings of drawings by 36 subjects and their scores on the Rosenberg Self-Esteem Scale. This led Dellate (1985) to conclude that "boys tend to project their feelings of self-esteem onto a figure drawing by varying its size, and girls tend to project self-esteem onto a figure drawing by varying its femininity" (p.166). What role elaborate details play in the projection of self-esteem is still not entirely clear.

While the meaning of elaborate details is relatively uncertain, it does seem that such details occur consistently across both human figure drawings and self-drawings. Short-DeGraff, Slansky, and Diamond (1989) compared the performance of 15 preschool age boys and girls on the Goodenough-Harris Draw-A-Person test, on a method of assessing self-drawings devised by Ayres and Reid (1966), and on various verbal subscales of the Wechsler Preschool

and Primary Scale of Intelligence (WPPSI). Both the Goodenough-Harris and the Ayres and Reid system are designed to measure the degree of realistic details present in drawings. With these scoring systems, Short-DeGraff et al. found that self drawings had a significant correlation with the Draw-A-Man test of .74 ($p < .05$), and a highly significant correlation with the Draw-A-Woman test of .91 ($p < .01$). Interestingly, none of the verbal subscales of the WPPSI were significantly related to any of the figure drawings, suggesting that drawings pull for performance rather than verbal abilities. In any case, it seems clear that there is some relationship between self-drawings and the traditional Draw-A-Person test. What remains unclear is how the level of realistic details in these drawings is related to self-esteem.

Sex Roles and Gender Issues

While many contemporary researchers have examined how the variable of self-esteem is projected into human figure drawings, several other researchers have investigated the ways in which the Draw-A-Person may reflect gender issues. Teglasi (1980) found that a woman's sex role orientation can influence the order in which she draws male and female figures on the Draw-A-Person. Teglasi administered the Draw-A-Person and the Wellesley Sex-Role Orientation Scale to 150 female undergraduates, and found that women who had a high traditional sex role orientation score were

significantly more likely to draw a male figure first than were women who had a low traditional sex role orientation score. In a follow up experiment, Teglasi compared 40 married women who were members of the National Organization for Women, with 80 married women who were not members of the National Organization for Women, and who were recruited simply by going door to door in rural Pennsylvania. None of the women who were members of NOW drew a male figure first, while 26% of the non-members drew a male figure first. This difference was statistically significant. Teglasi concludes that while the Draw-A-Person was designed to measure individual personality differences, it may also reflect "some broader cultural factors, such as attitudes toward sex roles" (p.271).

Rierdan, Koff, and Heller (1982) also obtained results with the Draw-A-Person which they believed were influenced by the sex role orientation of their subjects. Rierdan et al. obtained human figure drawings from a normal population of males and females, who were between the ages of 9 and 22. The drawings were scored according to the indices devised by Saarni and Azara (1977), for "anxiety related to aggression-hostility" and "anxiety related to insecurity-lability" (Rierdan et al., p.594). Male subjects evidenced more anxiety about aggression-hostility than did females, and both male and female subjects drew male figures that possessed more aggression-hostility indices than did the

female figures. Rierdan et al. explained these results by suggesting that males in our society are socialized to have more concern about aggression, and that both males and females are encouraged to think about males as more aggressive.

While Reirdan, Koff, and Heller (1982) explored how the drawings of normal subjects are affected by gender issues, Zucker, Finnegan, Doering, and Bradley (1983) investigated the ways in which drawing performance may be affected by the presence of gender identity disturbance. This research team obtained Draw-A-Person protocols from 36 children who were referred to a psychiatric institute due to potential problems in their gender identity development. These children were judged by intake workers to meet the DSM-III criteria for gender-identity disorder of childhood. Children from a sibling group, a psychiatric group, and a normal group all served as controls. Zucker et al. found that the gender-referred children were significantly more likely to draw an opposite sex figure first than were either the sibling, the psychiatric, or the normal controls. The gender-referred children who drew an opposite sex figure first were also significantly more likely to play with opposite sex toys and dress-up clothes in and unstructured play session, than were gender-referred children who drew a same sex figure first. The gender-referred also drew significantly taller opposite sex figures than same sex

figures. Finally, Zucker et al. also found that when all the drawings were scored for Koppitz (1968)'s criteria for emotional disturbance, the normal children had a significantly smaller proportion of psychopathology indicators than did the other three groups. In conclusion, Zucker et al. suggest that the Draw-A-Person test can be an effective way to assess gender identity disturbance in children.

Sexual Attitudes, Feelings, and Dysfunction

In the same way that an individual's figure drawings can tell us something about an individual's gender identity, empirical research also suggests that drawings can reveal something about a person's sexual functioning. While the diagnostic power of drawings may be limited, recent findings suggest that they can give us some information about both normal and pathological sexual functioning.

With 40 undergraduate subjects, Przybyla, Byrne, and Allgeier (1988) discovered that sexual attitudes correlate with the level of sexual details in human figure drawings. Przybyla et al. gave subjects the Sexual Opinion Survey, and asked them to draw nude human figures. Przybyla et al. report:

> that individuals with relatively positive sexual attitudes (erotophiles), as compared with individuals with relatively negative attitudes (erotophobes), were more likely to include such details as a glans, a urinary meatus, and chest hair on male figures and pubic hair on female

> figures. Positive sexual attitudes were also associated with drawing figures with longer and wider penises, breasts, testicles, and mons. (p.99)

On the basis of these significant findings, Przybyla et al. suggested that the draw-a-person in the nude technique might eventually be useful to clinicians who assess and treat sexual dysfunction. While their results were with undergraduates and must be viewed as tentative, they suggested that nude figure drawings can successfully represent one's attitudes about sexual behavior.

Miller, Veltkamp, and Janson (1987) were equally optimistic about the utility of projective drawings in the assessment and treatment of sexually abused children. Miller et al. suggested that projective drawings can help clinicians determine "the type of sexual abuse which the child has suffered, who the perpetrator is, and the victim's feelings" (p.51). These authors advocated drawings as a way of helping children express and begin to come to terms with feelings that are initially too difficult to discuss. Miller et al. suggest many useful instructions about how to elicit drawings from disturbed children, such as: "draw what your parents do when they are mad, draw a wish, draw a feeling, draw a daydream, etc..." (p.49). Several case studies were described which vividly illustrate the usefulness of projective drawings in the treatment of sexually abused children. Unfortunately, this anecdotal evidence is the only data that Miller et al. cite to support

their claims about the usefulness of projective drawings.

Yates, Beutler, and Crago (1985) took a somewhat more empirical approach to their study of drawings by child victims of incest than did Miller, Veltkamp, and Janson (1987). This research team compared 18 female incest victims at an outpatient clinic, to a control group of 17 patients who were also outpatients but were not incest victims. The control group was matched to the victim group by age and socioeconomic status. The drawings of both groups were rated by two clinical psychologists who were blind as to the subject's involvement in incest on "eighteen characteristics of potentially disturbed functioning that were extracted from the clinical literature" (Yates et al, p.185). Unfortunately, no mention is made of what criteria were used to determine or score the presence of these 18 characteristics in the figure drawings. Of the 18 characteristics, only two were significantly different between the two groups. The incest victims were judged to have more poorly developed impulse controls and greater use of repression as a defense mechanism. While these findings initially appear discouraging, they must be considered tentative because of the many methodological problems that plagued this study. Yates et al. reported that the drawings were all given by one physician who had no formal training in the administration of projective techniques, and who was aware of which subjects were incest victims. Yates et al.

also conceded that their sample size was inadequate, that they failed to score the drawings in a standard manner, and that their control group was not drawn from a normal population. Because of these problems, Yates et al. recommended that further research be done with a larger population, before projective drawings be abandoned as an assessment device for incest victims.

Sidun and Rosenthal (1987) completed a more controlled study of the Draw-A-Person protocols of psychiatrically hospitalized adolescents. Sidun and Rosenthal compared 30 adolescent inpatients with a previous history of sexual abuse, with 30 adolescent inpatients with no history of sexual abuse. The two groups were carefully matched for gender, age, IQ, race, and DSM-III diagnoses. The drawings of both groups were scored for a number of structural features. Results indicated that the abused adolescents were significantly more likely to omit hands, to draw figures with wedges, to draw phallic-like objects, and to draw figures with circles. Sidun and Rosenthal also report that there were trends for the abused group to omit fingers, and to draw pictures of only a head. While the presence of any one or combination of these graphic signs cannot conclusively diagnose a history of sexual abuse, Sidun and Rosenthal suggested that they can and should serve as warning signs to the astute clinician.

In a similar fashion, Johnston and Johnston (1986)

found several warning signs, when they compared the human figure drawings of 23 convicted child molesters with 28 college students. The two groups were matched for age and race. The child molesters produced figures of significantly poorer gender differentiation than did the college students, and the child molesters drew significantly more male figures with blank or missing eyes. The overall quality of the child molester's drawings of male figures was also found to be significantly poorer. As a group, the child molesters also produced male figures that were smaller than their female figures, with the difference in size being highly significant. No statistically significant difference was found between the size of the male and female figures that were produced by college students. Johnston and Johnston believed their findings may represent poor gender identity and low self-esteem on the part of the child molesters.

Other Psychological Variables

In addition to telling us about sexual dysfunction and deviancy, contemporary research suggests that the Draw-A-Person can help us explore a number of other psychological variables as well. Instead of selecting one sign or structural feature of the Draw-A-Person, and attempting to determine what personality traits correlate with this structural feature, contemporary researchers have tended to take the opposite approach. These scientists have chosen a particular personality trait or emotion, and then attempted

to determine if this psychological variable influences performance on the Draw-A-Person.

For instance, Daum (1983) compared both aggressive delinquents, withdrawn delinquents, and undifferentiated delinquent adolescents with a non-delinquent control group. A subject in the aggressive delinquent group "had to have a minimum of two contacts with the court for a hostile, aggressive crime, and it was necessary for his social history to report aggressive behavior independent of the court offenses" (p.245). For the withdrawn group, "an adolescent could have no court contacts for a hostile, aggressive crime but at least two charges of runaway or truancy. A withdrawn subject's social history had to mention shyness or fearfulness" (p.245). The aggressive delinquents drew significantly more figures with square shoulders than the other groups, while the withdrawn delinquents were significantly more likely to omit facial features, omit arms, and produce dimmer facial features than the other groups. While these indicators occurred relatively infrequently, Daum suggested that they can serve as warning signs to the clinician interested in human figure drawings.

Seifert (1988) has suggested that human figure drawings may also be useful to mental health professionals and educators who are interested in the assessment and treatment of autism. Seifert based this belief on her work with

Kenneth, an autistic adolescent. Seifert had the opportunity to assess Kenneth both before and after a four year placement in a residential school for severely autistic adolescents. While not empirical in nature, the discussion of Kenneth's case history provides thought-provoking anecdotal evidence about the usefulness of the Draw-A-Person. Seifert makes the point that drawings by autistic children "simultaneously tap one of the strengths (non-verbal ability) and one of the weaknesses (human-relatedness) of these children" (p.80). Drawings may offer us a unique window into both the strengths and weaknesses of the autistic child or adolescent.

While figure drawings may be quite useful with autistic children, Ginzburg, Merskey, and Lau (1988) have suggested that drawings are less useful when they are given to medical patients in pain clinics. The drawings given to pain patients, however, are very different from the traditional Draw-A-Person. Pain patients are presented with outlines of the human figure, and are asked to shade in the areas of the body where they feel pain. Ginzburg et al. studied pain patients drawn from four different settings: an anesthesiologist's pain clinic at a university teaching hospital, a dental clinic for facial pain, a psychiatric pain clinic, and an anesthesiologist's pain clinic at a rural general hospital. In addition to their pain drawings, these subjects were also given numerous psychological

questionnaires, to measure both their premorbid functioning and their current level of psychopathology. Ginzburg et al. found only a "very limited relationship between the extent of the body surface area affected by pain and the premorbid functioning and psychological state of subjects" (p.145). The study's strongest finding was that psychiatric patients tended to include a significantly larger area of the body in their pain drawings, than did patients from the two clinics directed by anesthesiologists. "The significance of this may be related to factors which are not necessarily psychological, for example, selection for nerve blocks or for psychiatric attention" (Ginzburg et al., p.145). In conclusion, Ginzburg et al. stated that:

> it seems inappropriate to rely on the amount of body area involved as any sort of proof that the patient has either a current psychological problem or a long-standing personality disorder. (p.145)

Shaffer, Pearson, Mead, and Thomas (1986), on the other hand, did find a meaningful relationship between the Draw-A-Person and the physical well-being of their subjects. This research team obtained figure drawings from 581 students at the John Hopkins University School of Medicine during the years of 1951-64. An extensive effort was then made to follow the physical and mental health of these subjects over the next two and a half decades. In 1984, subjects were placed into eleven health outcome categories: healthy (N = 386), suicide (N = 7), mental illness (N = 26), emotional

disturbance (N = 69), hypertension (N = 56), coronary occlusion (N = 9), other coronary disease (N = 7), major cancer (N = 28), skin cancer (N = 33), benign tumor (N = 83), duodenal ulcer (N = 34), and other death (N = 5). The drawings of all subjects were scored according to a sophistication-of-body-concept, and a conventional/deviancy scale.

While differences between groups in the sophistication-of-body-concept scores were generally not significant, a number of statistically significant differences did emerge with regards to the conventionality/deviance scale. This scale was originally developed by Thomas, Jones, and Ross (1968), and consists of 42 structural signs of conventionality/deviance. Three one-way ANOVAs were performed on the conventionality\deviance scale scores, and all three ANOVAs were statistically significant: (1) for all twelve groups, (2) with the healthy and other death groups omitted, and (3) with the other death group omitted. To follow up on these results, dichotomous group analyses were performed, in which all subjects in a specific outcome category were contrasted with all subjects not in that category. The healthy and coronary occlusion group means were both significantly higher than the means of all other groups, and their was a nearly significant trend for the duodenal ulcer group mean to be higher than all other groups. The mental illness and benign tumor group means were

both significantly lower than all other groups. In their discussion of these results, Shaffer et al. (1986) suggested that information obtained in figure drawings may be related to subsequent health status. These authors warned the reader, however,

> that the relationships found explained comparatively little of the total variance involved; thus, the degree to which a figure drawing measure of conventionality/ deviancy would be of practical value in predicting later health status is uncertain at the present time. (p.368)

With this warning statement, Shaffer et al. (1986) could easily have been describing the entire field of contemporary Draw-A-Person research. Most of the studies discussed that attempted to relate performance on the Draw-A-Person to specific emotions, personality traits, or aspects of individual psychopathology (such as aggression, autism, predisposition to illness, etc.), obtained some significant results. Unfortunately, the actual utility of these results for practicing clinicians has yet to be determined. Most authors conclude by suggesting that they have discovered warning signs, rather than conclusive indicators of particular traits (Daum, 1983; Johnston & Johnston, 1986; Shaffer et al., 1986; Sidun & Rosenthal, 1987). The astute reader of the current literature should also remember the advice of Basow (1986), who has suggested that referees for scholarly journals have a bias towards studies that reject the null hypotheses, at the expense of studies that fail to reject the null hypothesis. With these

caveats in mind, a modest conclusion would perhaps be that the Draw-A-Person has some validity with particular populations. The findings of current researchers suggest that further scientific exploration is warranted to refine our diagnostic knowledge about the Draw-A-Person test.

Cultural and Environmental Factors

In addition to assessing various emotions and individual characteristics, contemporary research suggests that the Draw-A-Person can also reflect an individual's upbringing, present environment, and cultural values. The test is not merely a reflection of individual personality, but is also a reflection of the subject's social world.

Poster, Betz, McKenna, and Mossar (1986) advanced this idea by demonstrating how human figure drawings can reflect the attitudes of children towards the mentally ill. Poster et al. asked 168 children in grades three through six to draw pictures of individuals who were "normal," as well as individuals who were "crazy" (p.680). Poster et al. report that:

> work/chores and play were predominant themes in drawings and stories depicting 'normal' behavior, while inappropriate behavior, suicide, aggression/hostility and self-abusive behavior were predominant themes in drawings and stories depicting 'crazy' people." (p.680)

In their discussion of the results, Poster et al. provide vivid examples of the violent behaviors which children attribute to the mentally ill. Children described crazy

individuals not merely as eccentric, but as people who shot innocent bystanders or repeatedly stabbed themselves. Poster et al. suggest that "these themes may be in part the result of children's exposure to increased discussion and media portrayal of rape, family violence, suicide, and other forms of inappropriate behavior" (p.685). If this explanation is correct, then educational efforts may be needed to help children overcome their misconceptions about mental illness. The results obtained by Poster et al. suggest that projective drawings could be an effective way to assess the success or failure of such an educational program.

Human figure drawings can also reveal some of the ways in which a child has been affected by his or her socioeconomic status. Pfeffer and Olowu (1986) investigated the effects of socioeconomic differences on the sophistication of Nigerian children's human figure drawings. Their subjects were 125 Yoruba school children from middle and low income schools. Children from the middle income school generally drew figures that were more realistic. The middle class children drew figures which had a more conventional shape, were more likely to contain all body parts, were more likely to have body parts in the correct position, and were more often clothed, than did the lower class children. All four of these differences were statistically significant at a high level. Unfortunately,

pfeffer and Olowu apparently made no attempt to control for the intelligence of their subjects. It is therefore not clear if the obtained differences were related to socioeconomic status, to some difference in the intellectual level of the two groups, or to a combination of both factors. Goodenough and Harris (1963) reported that children with higher IQ's also drew more complete and realistic figures. A task for future researchers might be to understand the ways in which both socioeconomic status and intelligence may simultaneously affect performance on the Draw-A-Person.

Future investigators may also need to consider how cultural factors can influence performance on the Draw-A-Person, given the results of two research teams that studied human figure drawings from a cross-cultural perspective. Koppitz and Casullo (1983) compared the human figure drawings of 147 Argentine adolescents with 147 USA adolescents. The two groups were matched for sex and exact age in years and months. Both samples were taken from a cross section of predominantly white, lower to middle-class families. The drawings were scored with the developmental scoring system of Koppitz (1966), as well as with a system of emotional indicators devised by Koppitz (1982). A number of significant differences were found between the two groups. Koppitz and Casullo report that:

> the Argentine youngsters, as a group, were better controlled, less aggressive, more evasive, and more concerned with appearance and action. The drawings of the USA pupils, as a group, displayed more often tendencies to be outgoing, impulsive, insecure, and aggressive. Cultural influences were also shown in the presence of "masculine" and "feminine" items, with the Argentine youngsters producing drawings with more gender characteristics. (p.479)

In a similar fashion, Munroe and Munroe (1983) also discovered several significant cross-cultural differences. Munroe and Munroe compared three different cultural groups in Kenya. Two of the groups, the Kipsigis and the Logoli, were highly in favor of moderninzation, while the third, the Gusii, continued to favor more traditional ways. Munroe and Munroe hoped to test the "values" hypothesis of Dennis (1966), who suggested that "whatever their own dress, or the dress of their community.... children most often draw the costume they admire" (Dennis, p.46). After collecting over 300 drawings, Munroe and Munroe scored the figures for the modern/tradtional nature of their dress, and obtained an interrater reliability coefficient of .87. For male figure drawings, it was found that the two more modern tribes (the Kipsigis and Logoli) produced significantly more modern drawings than did the more traditional tribe (the Gusii), and that male subjects, who were more involved in modernization, produced significantly more modern figures than did female subjects. For female figure drawings, no statistically significant results were obtained. Munroe and

Munroe suggest that all three tribal groups may have considered females to be less involved in the modernization process, and therefore drew relatively traditional female figures. In any case, the results of Munroe and Munroe suggest the rich potential of figure drawings to assess cross-cultural differences.

In many ways, this potential is still being discovered. Contemporary researchers have expanded our knowledge of the psychological variables that can be assessed with the Draw-A-Person. In most cases, however, the results are equivocal and tentative. The Draw-A-Person appears to warrant further research, in terms of both individual and cultural variables.

Structural and Formal Aspects of the Draw-A-Person

Most of the contemporary researchers cited above were interested in a particular psychological variable, and attempted to see how this variable might manifest itself on Draw-A-Person protocols. In the past, many researchers organized their work in the opposite fashion. These investigators began by focusing on structural aspects of drawings, such as size, or on a specific aspect of content, such as the nose, and attempted to see what psychological variables might be related to these particular signs. This approach may have been so popular due to the fact that Machover's (1949) influential manual was organized around a number of signs which dealt with both the structure and the

content of figure drawings. The three major published reviews of the empirical literature on the Draw-A-Person test were also organized around signs (Kahill, 1984; Roback, 1968; Swensen, 1968). A careful reading of these reviews suggests that the sign approach, though it was explored by hundreds of authors, was not terribly fruitful. Most of Machover's hypotheses about the content, as well as the structural and formal aspects of the Draw-A-Person, were not supported by empirical investigations. After these discouraging findings, it appears that contemporary researchers have abandoned projects that were designed to validate the usefulness of a particular sign. Since the time of Kahill (1984)'s survey, it appears that only size, shading, and the omission of eyes have been the focus of published investigations. Given the discouraging nature of most research on the structural and formal aspects of the Draw-A-Person, Machover's hypotheses and the conclusions of Roback (1968), Swensen (1968), and Kahill (1984) will only be reviewed briefly, along with a discussion of recent findings on size, shading, and the omission of eyes.

Ambiguously Sexed Figures

Machover (1949) believed that individuals who drew ambiguously sexed figures suffered from sexual maladjustment. Hammer (1954, 1958) believed that ambiguously sexed figures were indicative of homosexuality. The empirical evidence to support these hypotheses,

according to Kahill (1984), is mixed. A number of studies have found that homosexual males and females are more likely to draw ambiguously sexed figures than are heterosexual males and females (Kirchner & Marzoff, 1974; Pustel, Sternlicht, & Deutsch, 1971). An ambiguously sexed figure drawing, however, cannot be thought of as a clear indicator of homosexuality. Research suggests that it is common for heterosexual subjects to draw ambiguously sexed figures. Soccolich and Wysocki (1967) reported that in a sample of heterosexual college students, 43% of the male and 30% of the females drew ambiguously sexed figures. The only difference between heterosexual and homosexual subjects then, may be that it is slightly more common for homosexual subjects to draw ambiguously sexed figures.

Breasts

Machover (1949) stated that:

> the most consistent and significant interest in breast treatment is noted in the drawings of emotionally and psychosexually immature males. The breasts are erased, shaded, and returned to frequently for some additional furtive lines to mark preoccupation with that part of the figure. (p.69)

For female subjects, Machover (1949) believed that breast emphasis indicated a strong identification with a dominant mother image. Hammer (1954), on the other hand, suggested that breast emphasis in women might be compensation for feelings of sexual inadequacy.

Empirical research has generally failed to support the

hypotheses of either Machover (1949) or Hammer (1954) about breast emphasis (Kahill, 1984; Roback, 1968). The one exception to this was the work of Holzberg and Wexler (1950). This research team found that schizophrenic males tended to draw significantly larger breasts than did a control group of normal males, perhaps offering vaque support for the idea that males who are somehow disturbed or emotionally troubled will emphasize breasts. Researchers have been unable, however, to link breast emphasis directly to psychosexual immaturity in males subjects.

For female subjects, the work of Reirdan and Koff (1980) suggests that breast emphasis may simply be a part of normal development. Reirdan and Koff found that 16% of normal pubertal girls explicitly represented breasts on their figures drawings, while only 7% of female college students did so. It may be that breast emphasis simply reflects the normal preoccupation of young adolescent girls with the physical changes that they are rapidly experiencing.

Contact Features

Machover (1949) believed that arms, hands, and fingers were "contact features" (p.59), and reflected the interaction of the person with his or her environment. Arms which were drawn well to the side of the figure were thought to represent greater interpersonal warmth and confidence in social interactions than did arms that were close to the

trunk of the figure. Missing hands were thought to represent lack of confidence in social situations and feeling of personal ineffectiveness. Machover specifically believed that anti-social individuals would be likely to draw figures with hands hidden in their pockets. In a widely used manaul on Draw-A-Person interpretation, Jolles (1964) made a different suggestion. Jolles hypothesized that hands in pockets suggested masterbatory guilt. Fingers were considered the most immediate contact points by Machover (1949) and Hammer (1954), and were thought to be related to social communication, manipulation of the environment, and aggression.

A number of researchers have investigated the meaning of contact features, and have obtained mixed results (Kahill, 1984; Roback, 1968). It is unclear whether or not individuals who lack social confidence and feelings of personal effectiveness are more likely to draw figures with distorted hands and fingers. According to Roback (1968) and Kahill (1984), some investigators have found signficant differences, while other research teams have not. It does seem fairly clear, however, that individuals with anti-social personalities are no more likely to draw figures with hands in their pockets than are normals (Craddick, 1962). Jolles' (1964) hypothesis, of a relationship between hands in pockets and guilt about masterbation, has never been empirically tested.

Detailing

Both Machover (1949) and Hammer (1958) suggested that overly-meticulous and excessive details connoted an obsessive personality style. Early researchers did not explore this hypotheses, and Roback (1968) and Swensen (1968) did not discuss it. Kahill (1984), however, reported that in regard to detailing, "findings are mixed and subject to various interpretations" (p.276). Both relaxed subjects and serious, highly motivated subjects seemed to produce drawings with many details. With such mixed results, the precise meaning, therefore, of detailing remains uncertain.

Distortion

Hammer (1954, 1958) believed that distortion indicated severe emotional upheaval, and that bizarre distortion could be considered a sign of schizophrenia.

The attempts to verify this hypothesis have been mixed. In a review of the early literature on this relationship, Handler and Reyher (1965) concluded that "a majority of studies report significant relationships between distortion and severe psychopathology" (p.313). In many of the studies reviewed by Handler and Reyher, however, initial diagnosis was the only assessment obtained of the subjects' level of psychopathology. More recent investigators, who have used more psychometrically based methods of determining the subject's level of psychopathology, have had less encouraging results. Kahill (1984) suggested that the

current evidence is largely negative. While some studies do suggest a connection between distortion and psychotic thinking, other investigators have also linked highly distorted drawings with high creativity (Schaefer, 1982) and with immature cognitive skills (Kay, 1978). These findings suggest that the precise meaning of distortion remains unknown.

Ears

The ears, like the eyes, were thought to be an organ of contact with the outside world, and their emphasis was again thought by Machover (1949) and Hammer (1954) to suggest paranoia. Roback (1964) and Swensen both concluded, however, that no evidence exists to support this hypotheses. Numerous studies compared the treatment of ears by paranoid and non-paranoid subjects, and failed to discover any significant differences.

Erasure

Erasure was also considered an expression of anxiety by both Machover (1949) and Hammer (1958). When discussing erasure, Machover stated that "this form of conflict treatment is seen mostly in neurotics, obsessive-compulsive characters, and in psychopaths with neurotic conflicts" (p.98). The evidence regarding erasure and anxiety is clearly negative. Roback (1968), Swensen (1968), and Kahill (1984) all indicate that no studies have discovered a significant relationship between these two variables.

Kahill states that when researchers compared the erasure produced by highly anxious subjects with the erasure produced by relaxed subjects, they failed to find any significant differences. It remains unclear what, if anything, is indicated by the presence of erasure.

Eyes

Machover (1949) considered the eyes to be "a basic organ for contact with the outside world" (p.49). She believed that the eyes are emphasized by suspicious, paranoid individuals, who are searching for danger in the outside world. Hammer (1954) suggested that large, emphasized eyes could represent sensitivity, or the possible presence of visual hallucinations. Both Machover and Hammer hypothesized that eyes without pupils indicated immature, self-absorbed, or schizoid individuals.

Roback (1968) and Kahill (1984) both concluded that the empirical evidence is mixed, with a slight majority of studies failing to find a relationship between emphasized eyes and paranoia or suspicion. There is apparently no evidence to support Hammer's (1954) hypothesis that large eyes represent the presence of visual hallucinations.

There is some limited support for the notion that eyes without pupils represent immaturity. Wysocki and Wysocki (1977) found that 58% of incest offenders drew eyes without pupils, compared to 28% of child molesters, and only 8% of convicted rapists. Wysocki and Wysocki note that this

finding is consistent with the prevalent hypothesis that a cause of incest and child molesting is a narcissistic identification with immature objects. Kurdek and Darnell-Goetschel (1987), on the other hand, found in their sample of 44 normal adolescents that the omission of eyes was significantly related to high anxiety scores on the Symptom Checklist-90-R. Given these findings, it remains unclear what exacly is suggested by the omission of pupils or eyes.

Face

Machover (1949) considered the face to be the social feature of the drawing. She thought that omitted facial features indicated evasiveness, avoidance of social problems, superficiality, caution, and hostility in social contacts. Both psychopathic and paranoid individuals were thus expected to omit facial features more than normal individuals. To date, only a small amount of research has been conducted to test these hypotheses. Roback (1968) and Kahill (1984) report on several studies each, and conclude that there is currently no evidence to support Machover's ideas about the face.

Hair

Hair emphasis was regarded by Machover (1949), Hammer (1954, 1958), and Jolles (1964) as evidence of a desire to appear sexually potent and virule, possibly as compensation for feelings of sexual inadequacy or conflict. Machover (1949) also thought that immature subjects would be more

likely than mature subjects to draw figures with messy hair. Empirical evidence is generally unsupportive of the connection betwen hair emphasis and virility (Roback, 1968; Kahill, 1984). There does appear to be some connection, however, between severe psychopathology and disheveled or missing hair. Several studies have found that severely disturbed inpatients were significantly more likely to omit hair (Hozier, 1959), or to draw figures with disheveled hair (Cramer-Azima, 1956; Holzberg and Wexler, 1950) than were normals.

Inanimate Props and Special Themes

Hammer (1958) hypothesized that props, such as guns or knives, and soldier or cowboy themes, represented aggressive impulses, and would occur most frequently in the drawings of juvenile deliquents.

Only one study has attempted to test this theory directly. Montaque and Prytula (1975) found no differences between 30 normal and 30 deliquent adolescents in the incidence of either props or themes. The results of Montaque and Prytula cannot be considered conclusive, however, because none of their subjects, in either group, drew a figure with a prop or aggressive theme. Inanimate props and special themes are quite rare, but could possibly be an indication of aggression on the rare occasions that they do occur.

Mouth and Teeth

Machover (1949) believed that young children and primitive, regressed, or alcoholic individuals would all emphasize the mouth, as a demonstration of their unmet oral needs. Jolles (1964) also suggested that an emphasized mouth suggested immaturity and unresolved oral needs. A mouth with a heavy slash line, or a mouth with teeth showing, were both thought to be signs of aggression. Hammer (1954) also believed that teeth represented aggressive and hostile tendencies.

In an early review, Swensen (1957) indicated that the empirical investigations of the mouth and teeth had obtained mixed results, and that further research was warranted. Unfortunately, it appears that researchers failed to heed this recommendation, since both Roback (1968) and Swensen (1968) indicated that no new work has been done in this area. Kahill (1984), however, reported that some new work was later done on the meaning of the mouth and teeth, with mixed results. According to Kahill, a number of researchers have been able to link the presence of large teeth with aggression. Some investigators have also found that drug dependent subjects drew figures with larger mouths than did non-dependent subjects, while other scientists failed to discover any significant differences.

Nudity/Clothing

Both Machover (1949) and Hammer (1958) suggested that a scantily clad figure indicates a "body narcissist" (Hammer, p.265), who is likely to be schizoid and self-absorbed. Complete nudity, except when drawn by individuals such as artists, was thought to represent sexual maladjustment. The empirical research, however, suggests that nudity may be more of an indicator of gross psychopathology, than of narcissism. Researchers have found that psychotic patients were more likely to draw naked figures than were narcissistic patients or patients with other characterological problems (Kahill, 1984). It may be that psychotic individuals, who are experiencing a complete breakdown of their defenses and coping mechanisms, are the most likely to draw naked figures. The manner in which clothing may represent an individual's psychic defense mechanisms appears to warrant further investigation.

Placement

Machover (1949) suggested that a figure that is placed on the right side of the page indicates a subject who is environmentally-oriented, while a figure placed on the left side of the page indicates a subject who is self-focused. A figure placed high on the page was thought to suggest optimism, while a figure placed low on the page was thought to represent pessimism. Hammer (1958), on the other hand, thought that left side placement might indicate impusivity

and the need to seek immediate gratification. Unfortunately, none of these hypotheses appears to have been supported by empirical investigations. Roback (1968) concluded that investigators have failed to link placement with any other personality trait in a convincing manner, while Swensen (1968) and Kahill (1984) believed that the results were inconsistent, so that "for every study finding a significant relationship between placement and some behavioral characteristics, there exists a study relating similar kinds of data without significant results" (Swensen, p.31).

Sex of First Drawn Figure

Machover (1949) believed that it was normal for individuals to draw the same-sex figure first. She suggested that individuals who draw the opposite-sex figure first are likely to suffer from some degree of sexual inversion. Hammer (1954) also believed that individuals who are confused about their sex-role or sexual orientation would be more likely to draw the opposite- sex figure first, than would normal individuals. The hypothesis regarding the psychosexual significance of sex of first-drawn person has for the most part failed to find experimental support (Kahill, 1984; Roback, 1968). Investigators have also been unable to relate drawing the opposite-sex figure first to other forms of psychopathology, such as neurosis or drug abuse, in any consistent manner (Kahill, 1984).

There is considerable evidence, however, that it is relatively common in our society for women to draw the opposite-sex figure first (Daoud, 1976; Melikian, 1972; Soccolich & Wysocki, 1967; Teglasi, 1980), while males generally draw the same-sex figure first. Most researchers report that anywhere from 17% to 40% of female subjects will draw a male figure first. This may reflect remnants of sexism in our society, and a pervasive belief on the part of women that the male role is still more desirable. Paludi (1978) has suggested that our culture, under the influence of patriarchy, still equates being human with being male, and this causes women to draw a male figure first.

Size

Of the various structural aspects of the Draw-A-Person test, size has perhaps been considered the most extensively. Both Machover (1949) and Hammer (1958) hypothesized that size was related to energy level and self-esteem, with high energy and high self-esteem subjects tending to draw larger figures. Few investigators have attempted to study the relationship between size and energy level, perhaps because energy level is a difficult construct to operationalize and measure. Many investigators, on the other hand, have studied the relationship between size and self-esteem. In their review of the research literature, Swensen (1968), Roback (1968), and Kahill (1984) all concluded that the empirical evidence was mixed as to the relationship between

size and self-esteem. While some studies found that larger drawings are related to positive self-esteem, other investigators found no significant relationship. No study, however, has linked smaller drawings with higher self-esteem.

Since the publication of Kahill's review, investigators have continued to explore the meaning of the figure's size. Holmes and Wiederholt (1982) compared the Draw-A-Person protocols of depressed inpatients, non-depressed inpatients, and non-depressed hospital employees. The results of Holmes and Wiederhold suggest that depression and figure size are not related, since no statistically significant difference was found for figure drawing size between any of the three groups. Duffy, Beaty, and DeJulio (1982) on the other hand, did obtain significant results, when they asked 95 undergraduates "to draw a 'sexy' and an 'average' man, as well as a 'sexy' and an 'average' woman" (p.191). The sexy drawings were significantly larger than the average drawings, and subjects drew male figures that were significantly taller than the female figures. In the interpretation of their results, Duffy, Beaty, and DeJulio suggest that it may be normal for subjects to draw their sexual self larger than their normal self. They suggest that individuals who draw a sexy figure that is smaller than their regular figure may feel sexually inadequate.

Shading

Shading is the use of light lines to accentuate a particular part of the figure drawn. Machover (1949) and Hammer (1954) believed that shading is an indication of anxiety, and that the particular area shaded is the focus of the person's conflicts. For instance, figures that are shaded in the genital areas would suggest subjects with sexual concerns. Swensen (1968) believed that research has failed to support the relationship between anxiety and shading. Swensen indicates that researchers found no relationship, or else found that less anxious individuals produced figures with more shading. Swensen suggests that if shading is related to anything, it is related to drawing quality, with drawings of higher quality generally containing more shading. This might explain why less anxious subjects sometimes produce drawings with more shading, since less anxious, better adjusted subjects are generally thought to produce drawings of higher quality. Roback (1968) and Kahill (1984) on the other hand, both believed that the results of the existing research literature on shading are simply inconclusive.

A recent study, conducted by Kurdek and Darnell-Goetschel (1987) does suggest that shading is related to anxiety. Kurdek and Darnell-Goetschel administered the Draw-A-Person and the Symptom Checklist-90-R to a group of 44 adolescents who were drawn from a normal population.

Results indicated that high anxiety scores were significantly related to face shading, body shading, and hands shading. Based on these findings, Kurdek and Darnell-Goetschel suggest that shading can be a reliable measure of anxiety.

Transparency

A transparency is when a body part is showing through clothing, or an internal organ is showing through skin. Hammer (1954) believed that transperencies are a denial of reality and represent psychotic features. Handler (1967) regarded transperencies as signs of anxiety. Unfortunately, both Swensen (1968) and Kahill (1984) conclude that researchers have been unable to relate transpercies to any personality trait or emotion in a consistent manner.

Trunk

According to Machover (1949), round trunks tend to be drawn by passive individuals with feminine characteristics, while square trunks are thought to be drawn by masculine individuals. It appears that Janzen and Coe (1975) are the only researchers to date, who have tested this hypothesis. These authors studied a normal population of adult women, and reported that homosexual women drew significantly more square trunks than did heterosexual women. It is unclear, however, if this finding supports Machover's (1949) hypotheses, since there is no emperical evidence that the

homosexual women were more masculine than were the heterosexual women.

Artistic Ability: A Possible Confound

The study of structure and content on the Draw-A-Person test ultimately raises the issue of artistic ability. Many clinicians have wondered whether subjects draw well-proportioned and detailed figures because they are psychologically healthy, or if subjects do so because they are artistically talented.

Hammer (1958) dismissed the idea that drawing performance was primarily a product of artistic ability. Hammer pointed out that a brief visit to any art gallery will clearly demonstate that even artistically talented individuals produce vastly different works of art. Hammer felt it was safe to assume that these artistic differences were a reflection of the artists' individual personalities.

Other psychologists, however, were more concerned about the potential influence of artistic ability on Draw-A-Person performance. Most research studies in this area have found some relationship between drawing ability and the assessment of psychological adjustment on the Draw-A-Person (Kahill, 1984; Roback, 1968; Swensen,1968). For example, Feldman and Hunt (1958) found considerable overlap between those body parts of figure drawings that were rated the most difficult to draw by artists, and body parts that were most frequently selected by clinicians as indicative of emotional

disturbance. Feldman and Hunt had 65 undergraduate subjects draw nude human figures. Three clinicians then rated 25 body parts on each drawing for the presence or absence of psychological disturbance. A group of art teachers were then asked to rate the difficulty of drawing each of these 25 body parts on a five point scale. According to Feldman and Hunt, "a correlation of -.53 ($p < .01$) was obtained by correlating the average ratings of the art instructors with the average z scores of the clinical judges" (p.219). Those body parts which were rated as the most frequent signs of emotional disturbance were also those signs considered the most difficult to draw. A number of more contemporary research psychologists have also obtained similar results, suggesting that at times, lack of artistic talent may be falsely interpreted as maladjustment (Cressen, 1975; Johnson & Greenberg, 1978; Solar, Bruehl, & Kovacs, 1970).

After reviewing the existing research literature on the relationship between artistic ability and the interpretation of human figure drawings, Feher, Vandecreek, and Teglasi (1983) reach two conclusions. First, these authors surmise that "clinician do rely heavily on art quality in their evaluation of human figure drawings" (p.274). At the same time, however, Feher et al. also conclude that human figure drawings continue to have some validity as tools of personality assessment. Drawings can continue to be useful, if clinicians are aware of how drawing ability can affect

the final product. In particular, it is essential for clinicians to have some awareness of which body features are the most difficult to draw. This will prevent practicing psychologists from incorrectly conluding that a strangely shaped body part is a representation of psychopathology, when it really indicates little more than a lack of artistic talent.

Draw-A-Person Performance and Learning Disabilities

In the same way that artistic ability can partially determine a subject's performance on the Draw-A-Person, research results suggest that the presence or absence of learning disabilities can affect the final Draw-A-Person protocol as well. Investigations on the relationship between the Draw-A-Person and learning disabilities have mostly taken place since the mid-1970's, when learning disabilities first became a national concern (Moses, 1990).

Ottenbacher, Abbot, Haley, and Watson (1983) studied 40 children between the ages of five and thirteeen who were all diagnosed by a multidisciplinary professional team as learning disabled. Human figure drawings were obtained from these children, and were evaluated according to the scoring system of Ayres and Reid (1966). The Ayres and Reid system is designed to measure the level of realistic details and perceptual accuracy that are present in a child's drawings. The variables of age, and sex were recorded, and an assessment was made of each child's IQ. The Southern

California Postrotary Nystagmus Test was given to each child, to obtain an estimate of their postrotary nystagmus duration (PRN). The Southern California Postrotary Nystagmus Test is thought to provide a direct meaure of each child's vestibular-ocular reflexes. A regression analysis then revealed that a signficant amount of variance in human figure drawing scores was shared first by chronological age, and then by chronological age and postrotary nystagmus durations. The variables of IQ and sex were not significant. These results suggest that

> learning-disabled children can be differentiated based on their performance on vestibular-ocular function, and that human figure drawings may be used as one measure of vestibular related dysfunction.
> (Ottenbacher et al, p.1087)

Prewett, Bardos, and Naglieri (1988) obtained somewhat more discouraging results when they compared the Matrix Analogies Test-Short Form (MAT-SF) to the Quantitative Scoring System for the Draw-A-Person developed by Naglieri (1986), as methods of screening students with possible learning disabilities. The MAT-SF is a screening test comprised of 34 items of the progressive matrix type. Forty-four regular and 33 LD fourth and fifth-grade students were given the MAT-SF and the Draw-A-Person, as well as the Kaufman Test of Educational Achievement. Both subject groups scored within the average range on the two screening tests, while the LD group scored significantly lower on the

achievement test than on the two screening tests. Prewett et al. also report that:

> The MAT-SF was found to correlate significantly with all areas of achievement for the normal group. The DAP did not correlate significantly with any areas for the normal group; it correlated significnatly with reading, but not with math, for the learning-disabled students. (p.352)

These results suggest that the MAT-SF may be a more useful screening device than the draw a person for detecting the presence of learning disabilities.

On the other hand, Cox and Howarth (1989) have suggested that the Draw-A-Person may be a useful way of studying the deficits of learning disabled children. Cox and Howarth studied three groups of 15 children each: a normal group of four year olds, a normal group of nine year olds, and a learning disabled group of nine year olds. All of these children were asked to draw a man, draw the arms on a series of incomplete figures, and copy lines in four different orientations. When the developmental quality of these drawings was assessed for body proportion and completeness, Cox and Howarth report that "the differences between the normal nine year olds and the other two groups were statistically significant, whereas the differences between the normal four year olds and the learning disabled nine year olds were not" (p.338). These results suggest that the responses of learning disabled children may represent a developmental delay rather than a disorder of

the complex skills involved in drawing. In any case, Cox and Howarth claim that the Draw-A-Person can be a useful research tool, as psychologists struggle to understand the nature of learning disabilities.

The Scoring of Cognitive Ability

Whether a researcher is attempting to study learning disabilities, or any other psychological variable with the Draw-A-Person, a reliable scoring system is an absolute necessity. Without some method of scoring the information contained in drawings, Draw-A-Person research will never be able to proceed in an empirical fashion. A number of psychologists have attempted to address this problem by developing and validating scoring systems.

To date, many of the available systems have focused on using the Draw-A-Person to measure cognitive ability, both with children and with adults. Since practicing clinicians often need a rapid, non-verbal method of measuring intellectual development, the potential use of the Draw-A-Person in the assessment of cognitive ability has been investigated widely. While individual intelligence tests, such as the Wechsler Adult Intelligence Scale-Revised, are generally considered the preferred method of assessing cognitive development, the administration of these tests is not always feasible. Individual intelligence tests are expensive and time consuming to administer, and many subjects are either unable or unwilling to take one of these

figures which are filled with such elaborate details as shoelaces, shirt collars, and eyebrows. While Goodenough concentrated on the intellectual component of children's drawings, she also recognized that the drawings revealed emotional maturity and psychopathology (Taylor, 1977).

Goodenough developed her cognitive scoring system by studying the drawings of 100 children who attended kindergarten through the fourth grade, in order to determine what features one could expect to obtain from each grade level. The system that Goodenough (1926) published was apparently quite reliable. According to Harris (1963):

> "a number of studies established the consistency with which scorers can, with a minimum of training, score the Goodenough Draw-a-Man Test. Intercorrelations between different scorings range from the low .80's to as high as .96. Values commonly exceed .90." (p.90)

Correlations of above .90 were common for inter-rater, test-retest, and split-half reliability coefficients (Harris, 1963). Research on the validity of Goodenough's system obtained somewhat more variable results. A number of studies compared scores obtained by subjects on Goodenough's system with the scores of these subjects on the Stanford-Binet Intelligence Test and the Wechsler Intelligence Scale for Children. In a review of this body of research, Harris indicated that for the Stanford-Binet, correlations ranged from .26 to .92, while for the Wechsler correlations ranged from .38 to .77, with the majority of the coefficients being

statistically significant at or beyond the .05 level. In general, research conducted in the United States tended to support the validity of Goodenough's system. Goodenough's original study was also successfully replicated in Europe, Africa, and Japan (Harris, 1963). While Goodenough's system was clearly less than perfect, it did prove to be very useful to clinicians in underdeveloped parts of the world, "where convenient nonverbal measures of intelligence were needed, to classify large numbers of non-reading children for educational purposes" (Harris, 1963, p.11).

The world-wide response given to her scoring system was encouraging enough to prompt Goodenough and Harris (1963) to undertake a major revision and normative study of Goodenough's (1926) original scoring method. This revision was designed to extend the scale upward into the adolescent years, and explore new items which might increase the reliability and validity of the scale. For their normative group, Goodenough and Harris obtained a sample of both rural and urban children from Minnesota and Wisconsin who were quite representative of the general population in terms of socio-economic status. More than 300 children were tested at each grade level from kindergarten through ninth grade. Unfortunately, these subjects were almost exclusively of western European origin, and were not representative of the general American population. Nevertheless, the sample obtained by Goodenough and Harris remains the most

impressive normative group obtained by any researcher studying the Draw-A-Person. From this data base, Goodenough and Harris obtained a reliable scoring system of 71 items with norms for children aged three to fifteen.

It is surprising, however, that no one has attempted to study the usefulness of the Goodenough-Harris system with adults. While the test was developed for the assessment of children, it may be the case that the intelligence of adults will also be reflected in the level of realistic details that are present in their drawings. There has been some research suggesting that adults who are better educated and more affluent will draw more detailed figures (Adler, 1971). Given the encouraging research that has been done on the reliability and validity of the Goodenough-Harris system with children, it seems worthwhile to test the system with adult subjects.

It would also be helpful if researchers would examine the scoring system that has traditionally been used to assess the cognitive level of human figure drawings by adults. Most researchers continue to use the H-T-P Technique (House-Tree-Person) that was developed by Buck (1948). This system is an attempt to measure the cognitive skills of adults by assessing the details, proportion, and perspective of each's subject's house, tree, and person drawing. The test was developed by using a sample of 140 adults of seven different intelligence levels. The seven

levels ranged from profoundly retarded to superior intellectual functioning, with 20 subjects in each level. The subjects of average and superior functioning were undergraduates and graduate students, while the borderline to retarded subjects were residents of a state home for the mentally retarded in Virginia. Retarded subjects were classified into the various intelligence levels on the basis of their long-term functioning and adaptation rather than on standardized test scores. An effort was made to screen subjects with emotional problems out of the study so that the drawings would reflect cognitive ability rather than psychopathology. According to Buck (1948):

> the 140 sets of drawings obtained were subjected to minute and careful analysis in an attempt to identify and list as many as possible of the items which might by their presence or absence serve to differentiate subjects on the basis of intelligence. (p. 7)

Buck eventually developed a system of approximately 40 points for scoring a human figure drawing. The system is somewhat unique in that it is based on both the presence and absence of a variety of signs, rather than on the inclusion of details.

Buck (1966) later published a revised manual for the H-T-P which included research suggesting that his H-T-P technique correlated well with standard measures of intelligence. In a study of 100 Caucasian adults at a state home for the mentally retarded, Buck (1966) obtained the following Pearson correlation coefficients for the H-T-P

technique with the Wechsler-Bellevue Intelligence Scale: for the H-T-P and Verbal IQ, r = . 699, for the H-T-P and Performance IQ, r = .724, and for the H-T-P and Full Scale IQ, r = .746 (the level of statistical significance for these coefficients was not reported). These findings were apparently conclusive enough to discourage further research into the relationship between the H-T-P and individual intelligence tests. Kahill (1984) did not cite any new research on this relationship in her exhaustive review of the published literature on the Draw-A-Person, and a review of the most recent literature suggests that nothing has been published since the time of Kahill's report.

The dearth of current research leaves several questions open for future exploration. First, researchers must determine if the person component of Buck's (1948) system relates to overall IQ, when house and tree drawings are not present. Second, it would also be useful to examine the utility of the Goodenough-Harris Drawing Test with adults. If a relationship is found between either Buck's scoring method and/or the Goodenough-Harris scoring method and overall IQ, then researchers should determine what components of IQ are most closely related to the Draw-A-Person scoring systems. No one has attempted a regression analysis on such data, which could determine which subtests or group of subtests on an individual IQ test can best predict performance on either Buck's system or the

Goodenough-Harris system. For instance, field independence, which Witkin (1965) has found to be a factor on the Wechsler Adult Intelligence Scale, might well be highly related to the ability to do a detailed drawing. Field independence, which is made up of the Picture Completion, Block Design, and Object Assembly subtests, is thought to measure an individual's ability to impose structure where it is lacking. This ability could be quite useful when subjects are given a blank piece of paper and told that they must shape an imaginary person. Researchers should explore whether or not the factor of field independence is more stongly related to cognitive performance on the Draw-A-Person than is Full Scale IQ. Finally, it is curious that no one has attempted to compare the usefulness of Buck's system and the Goodenough-Harris Drawing Test with adult subjects. While slight differences do exist between the two systems, there is also a great deal of overlap in the items they cover. A comparison of the two methods on one population seems logical.

Hypotheses

A review of the existing literature about the intellectual evaluation of human figure drawings raised a number of concerns about this body of information. In an attempt to address the concerns that were raised and further our understanding of human figure drawings, the following related hypotheses were tested:

1. In keeping with the limited amount of past research, it was predicted that scores on Buck's (1948) scoring system would be significantly and positively correlated with Verbal, Performance, and Full Scale IQ scores.

2. Since the Goodenough-Harris Drawing Test is similar in many ways to Buck's (1948) scoring system, it was also expected that scores on the Goodenough-Harris Drawing Test would be significantly and positively correlated with Verbal, Performance, and Full Scale IQ scores.

3. Given the large number of features that overlap between the Goodenough-Harris Drawing Test and Buck's (1948) system, it was predicted that these tests would be related in a similar fashion to subjects' performance on a standard, individually administered IQ test. In other words, it was predicted that the correlation obtained between the Goodenough-Harris Drawing Test and Performance IQ, Verbal IQ, and Full Scale IQ scores, would not be significantly different from the respective correlations obtained between Buck's (1948) system and Performance IQ, Verbal IQ, and Full Scale IQ scores.

4. It was expected, since the Draw-A-Person is a largely non-verbal measure, that Buck's (1948) scoring system and the Goodenough-Harris drawing test would both be

significantly more correlated with Performance IQ scores, than with either Full Scale or Verbal IQ scores.

5. The Draw-A-Person test requires that subjects take a blank piece of paper and form a figure with few environmental cues to guide them. For this reason, it was predicted that the factor of field independence, developed by Witkin (1965), and measured by the Picture Completion, Block Design, and Object Assembly subtests, would serve be more highly related to cognitive scores on the Draw-A-Person test, than would Verbal, Performance, or Full Scale IQ scores.

CHAPTER III

METHOD

Subjects

The test protocols of 101 subjects were selected for this study from the archives of the Loyola University Department of Psychology's assessment laboratory. The archives of the assessment laboratory consist of the records of undergraduates at Loyola University Chicago, who have volunteered to take a battery of standard psychological tests. These undergraduate subjects received partial credit towards the requirements of an introductory psychology course for their participation in the laboratory.

Since part of purpose of the laboratory is to train doctoral students in the administration of standard psychological tests, the protocols of the current study were administered by doctoral students in clinical psychology, between the years of 1988 and 1990. These students worked under the supervision of an experienced clinical psychologist. An effort was made to obtain protocols which doctoral students gave later in their training, after they had mastered the administration of the tests in question. Otherwise, the subjects were selected at random from the

several hundred cases available in the archives of the assessment laboratory.

Measures

The following materials were examined: 1) the first figure drawing of each subject's Draw-A-Person test, 2) the first human figure drawing of each subject's House-Tree-Person, and 3) the Wechsler Adult Intelligence Scale-Revised (WAIS-R) protocol of each subject.

Procedure

When the protocols of 101 subjects had been obtained, the investigator scored the first human figure drawing of each subject according to the Goodenough-Harris Drawing Test, and with the person component of Buck's (1948) House-Tree-Person Technique. The Verbal IQ, Performance IQ, Full Scale IQ, and 11 subtest scores from the Wechsler Adult Intelligence Scale-Revised were also recorded for each subject.

The two scores of cognitive development from the human figure drawings were correlated with the Verbal IQ, Performance IQ, and Full Scale IQ scores from the WAIS-R. To see if the Goodenough-Harris scores of cognitive development correlated at a higher level with performance on the WAIS-R than did the H-T-P scores of cognitive development, t-tests, as recommended by Hosteling, were performed.

Scores were also obtained for the factor of field independence, by summing the scores of each subject on Picture Completion, Block Design, and Object Assembly. A stepwise mutiple regression analysis was performed to determine if field independence, or any single subtest or group of subtests from the WAIS-R, served as a superior predictor of performance on Buck (1948)'s scoring system or on the Goodenough and Harris (1963) scoring system. A formula proposed by Wherry, and first reported by Lord and Novick (1968), was used to remove the cumulative sampling error from the multiple correlation coeffients obtained during the regression analysis.

CHAPTER IV

RESULTS

The 101 human figure drawings in this study were scored by two raters (a Ph.D. candidate in clinical psychology and an undergraduate who worked under his supervision). In order to determine the percentage of inter-rater agreement, both raters independently scored a randomly selected group of 20 drawings with each of the two cognitive scoring systems. With Buck's (1948) system, the two raters scored 96.1 % of the items in the same direction. With the Goodenough and Harris (1963) system, the two raters scored 93.4 % of the items in the same direction. In keeping with the recommendation of Cohen (1960), Kappa coefficients were determined for each of the interrater agreement figures, to account for the number of possible responses for each item of the scoring system. These figures were also quite encouraging. With Buck's system Kappa = .927, while for the Goodenough and Harris system Kappa = .872. Since there was such high inter-rater agreement, the remaining drawings were divided between the two raters. Each drawing was then scored by only one of the two raters.

After the drawings were scored, mean scores and

standard deviations were determined for all of the human figure drawing scores and IQ scores. These results are reported in Table 1.

Hypothesis #1

The first hypothesis predicted that scores on Buck's (1948) scoring system would be significantly and positively correlated with Verbal, Performance, and Full Scale IQ scores. To test this hypothesis, two scores were initially obtained for each figure drawing with Buck's system: a raw score and a weighted score. The weighted scores were based on a formula by Buck (1948), which increases the value of certain items that Buck thought were most closely related to individual IQ scores. Standard scores were also determined on the basis of the raw scores, with the two being highly correlated (for Buck's raw score with Buck's standard score, $r = .964$, $p < .0001$). Buck's standard scores, like IQ scores, were devised to have a mean score of 100 and a standard deviation of approximately 15. Once the raw, weighted, and standard scores were available on Buck's system, Pearson correlation coefficients were determined for each of these scores with Verbal, Performance, and Full Scale IQ. These correlation coefficients are reported in Table 2. These analyses indicated that cognitive scores on Buck's system were significantly correlated with both Performance and Full Scale IQ scores, but were not significantly correlated with Verbal IQ scores. There was,

Table 1

Means and Standard Deviations for Human Figure Drawing Scores and WAIS-R Scores

Variable	Mean Score	Standard Deviation
Buck's (1948) raw score	.80	.12
Buck's (1948) weighted score	24.85	12.60
Buck's (1948) standard score	100.14	18.14
Goodenough-Harris raw score	42.72	7.50
Goodenough-Harris standard score	93.54	13.27
WAIS-R subtest scores and IQs		
Information	9.82	2.29
Digit Span	11.07	2.56
Vocabulary	10.53	2.31
Arithmetic	10.21	2.12
Comprehension	10.98	2.56
Similarities	10.82	2.30
Picture Completion	9.94	2.60
Picture Arrangement	10.46	2.75
Block Design	11.35	2.83

(continued)

Table 1 (continued)

Variable	Mean Score	Standard Deviation
Object Assembly	10.70	2.91
Digit Symbol	12.22	2.60
Verbal IQ	110.29	11.82
Performance IQ	108.98	14.16
Full Scale IQ	110.99	12.21

Table 2

Pearson Correlations Coefficients for All Subjects (n=101) for Buck's (1948) Scoring System with Wechsler IQs

	Buck's (1948) Scoring System		
	Raw Score	Weighted Score	Standard Score
Verbal IQ	.164 *	.181 *	.150
Performance IQ	.403 ****	.432 ****	.383 ****
Full Scale	.326 ****	.356 ****	.305 ***

* p. < .10
** p. < .05
*** p. < .01
**** p. < .001

however, a trend for the raw scores and weighted scores to be correlated with Verbal IQ scores.

Given the relationship that was found between Buck (1948)'s scoring system and both Performance and Full Scale IQ scores, direct difference *t*-tests for correlated observations were performed to determine if Buck's standard scores were an accurate reflection of either Performance or Full Scale IQ scores. The results of these *t*-tests are indicated in Table 3. These results suggest that while Buck's standard scores were significantly correlated with both Performance and Full Scale IQ scores, Buck's standard scores significantly underestimate both Performance and Full Scale IQ scores.

Hypothesis #2

The second hypothesis predicted that scores on the Goodenough-Harris Drawing Test would be significantly and positively correlated with Verbal, Performance, and Full Scale IQ scores. To test this hypothesis, raw scores were initially obtained for each figure drawing with the Goodenough-Harris system. Standard scores were then determined on the basis of the raw scores, with the two being highly correlated (for Goodenough-Harris raw scores with Goodenough-Harris standard scores, $r = .979$, $p < .0001$). Once both the raw and standard scores were available for the Goodenough-Harris Drawing Test, Pearson correlation coefficients were determined for both the raw

Table 3

Direct Difference t-tests for Correlated Observations for Buck's (1948) Standard Scores with IQ (n=101)

	Mean Score	sd	Mean Difference	t value	p
Buck Standard Score	100.14	18.14			
			-8.84	-4.87	< .001
Performance IQ Score	108.98	14.16			

	Mean Score	sd	Mean Difference	t value	p
Buck Standard Score	100.14	18.14			
			-10.86	-5.88	< .001
Full Scale IQ Score	110.99	12.21			

and standard scores with Verbal, Performance, and Full Scale IQ scores. These correlation coefficients are reported in Table 4. These analyses indicate that cognitive scores on the Goodenough-Harris Drawing Test were significantly correlated with both Performance IQ and Full Scale IQ, but were not significantly correlated with Verbal IQ.

Given the relationship that was found between the Goodenough-Harris Drawing Test scores and both Performance IQ and Full Scale IQ, direct difference *t*-tests for correlated observations were performed to determine if the Goodenough-Harris standard scores were an accurate reflection of either Performance IQ or Full Scale IQ. The results of these *t*-tests are indicated in Table 5. These results suggest that while the Goodenough-Harris standard scores are significantly correlated with both Performance IQ and Full Scale IQ, the Goodenough-Harris standard scores significantly underestimate both Verbal IQ and Full Scale IQ.

Since both the Goodenough-Harris standard scores and Buck's (1948)'s standard scores significantly underestimated both Performance and Full Scale IQ scores, it seems logical to ask if one of the two human figure drawing scores produces a superior estimate of Performance and Full Scale IQ scores. To address this question estimates of omega squared were determined, in keeping with the recommendation of Hays (1981), to estimate the percent of variance

Table 4

Pearson Correlations Coefficients for All Subjects (n=101) for Goodenough-Harris Scores with Wechsler IQs

	Goodenough-Harris Scores	
	Raw Score	Standard Score
Verbal IQ	.101	.110
Performance IQ	.415 ****	.369 ****
Full Scale IQ	.306 ***	.288 ***

* p. < .10
** p. < .05
*** p. < .01
**** p. < .001

Table 5

Direct Difference t-tests for Correlated Observations for Goodenough-Harris Standard Scores with IQ (n=101)

	Mean Score	sd	Mean Difference	t value	p
Goodenough-Harris Standard Score	93.53	13.27			
			-15.46	-10.06	< .001
Performance IQ Score	108.98	14.16			

	Mean Score	sd	Mean Difference	t value	p
Goodenough-Harris Standard Score	93.53	13.27			
			-17.46	-11.52	< .001
Full Scale IQ Score	110.99	12.21			

accounted for by the differences between the standard scores of the two human figure drawing systems and both Performance and Full Scale IQ scores. These figures are reported in Table 6.

Hypothesis #3

Given the large number of features that overlap between the Goodenough-Harris Drawing Test and Buck's (1948) system, the third hypothesis predicted that these tests would be related in a similar fashion to subjects' performance on a standard, individually administered IQ test. In other words, it was predicted that the correlations obtained between the Goodenough-Harris Drawing Test and Performance IQ, Verbal IQ, and Full Scale IQ, would not be significantly different from the respective correlations obtained between Buck (1948)'s system and Performance IQ, Verbal IQ, and Full Scale IQ.

The *t*-tests, which Hosteling (1940) has recommended for testing differences between two dependent correlation coefficients, were performed. The correlation coefficients obtained for the Goodenough-Harris Drawing Test raw and standard scores with Performance IQ, Verbal IQ, and Full Scale IQ were compared to the respective correlation coefficients obtained for Buck's (1948) raw, weighted, and standard scores with Performance IQ, Verbal IQ, and Full Scale IQ. The results of these analyses are reported in Tables 7 to 9. As predicted, the correlations found between

Table 6

Estimate of Omega Squared (w^2) for All Subjects (n=101) for Standard Scores with Performance and Full Scale IQ Scores

	w^2
Buck's (1948) standard score with Performance IQ	.184
Buck's (1948) standard score with Full Scale IQ	.250
Goodenough-Harris standard score with Performance IQ	.498
Goodenough-Harris standard score with Full Scale IQ	.566

Table 7

The t-test Values Obtained for the Difference in the Pearson Correlation Coefficients of the Goodenough-Harris Drawing Test with Performance IQ (n=101) and Buck's (1948) Scoring System with Performance IQ (n=101)

Dependent Correlation Coefficients	t	p
Buck's (1948) raw score with Performance IQ (r = .403) to Goodenough-Harris raw score with Performance IQ (r = .415)	0.183	NS
Buck's (1948) weighted score with Performance IQ (r = .432) to Goodenough-Harris raw score with Performance IQ (r = .415)	0.257	NS
Buck's (1948) weighted score with Performance IQ (r = .432) to Goodenough-Harris standard score with Performance IQ (r = .369)	0.880	NS
Buck's (1948) standard score with Performance IQ (r = .383) to Goodenough-Harris standard score with Performance IQ (r = .369)	0.191	NS

Table 8

The t-test Values Obtained for the Difference in the Pearson Correlation Coefficients of the Goodenough-Harris Drawing Test with Verbal IQ (n=101) and Buck's (1948) Scoring System with Verbal IQ (n=101)

Dependent Correlation Coefficients	t	p
Buck's (1948) raw score with Verbal IQ (r = .164) to Goodenough-Harris raw score with Verbal IQ (r = .101)	0.877	NS
Buck's (1948) weighted score with Verbal IQ (r = .181) to Goodenough-Harris raw score with Verbal IQ (r = .101)	1.070	NS
Buck's (1948) weighted score with Verbal IQ (r = .181) to Goodenough-Harris standard score with Verbal IQ (r = .110)	0.904	NS
Buck's (1948) standard score with Verbal IQ (r = .150) to Goodenough-Harris standard score with Verbal IQ (r = .110)	0.504	NS

Table 9

The t-test Values Obtained for the Difference in the Pearson Correlation Coefficients of the Goodenough-Harris Drawing Test with Full Scale IQ (n=101) and Buck's (1948) Scoring System with Full Scale IQ (n=101)

Dependent Correlation Coefficients	t	p
Buck's (1948) raw score with Full Scale IQ (r = .327) to Goodenough-Harris raw score with Full Scale IQ (r = .306)	0.307	NS
Buck's (1948) weighted score with Full Scale IQ (r = .356) to Goodenough-Harris raw score with Full Scale IQ (r = .306)	0.502	NS
Buck's (1948) weighted score with Full Scale IQ (r = .356) to Goodenough-Harris standard score with Full Scale IQ (r = .288)	0.913	NS
Buck's (1948) standard score with Full Scale IQ (r = .305) to Goodenough-Harris standard score with Full Scale IQ (r = .288)	0.223	NS

the Goodenough-Harris Drawing Test and Performance IQ, Verbal IQ, and Full Scale IQ were not significantly different from the respective correlations found between Buck's scoring system and Performance IQ, Verbal IQ, and Full Scale IQ.

Hypothesis #4

Since human figure drawings are a largely non-verbal task, the fourth hypothesis predicted that both Buck's (1948) scoring system and the Goodenough-Harris Drawing Test would be significantly more correlated with Performance IQ, than with either Verbal IQ or Full Scale IQ. In order to test this hypothesis, *t*-tests were performed to compare the Pearson correlation coefficients obtained between scores on both Buck's system and the Goodenough-Harris Drawing Test with Performance IQ, to the correlation coefficients obtained between these two scoring systems and both Verbal IQ and Full Scale IQ. The results of these analyses for Buck's system are reported in Table 10, while the results for the Goodenough-Harris Drawing Test are reported in Table 11.

For Buck's (1948) system, the correlation coefficients obtained with Performance IQ were significantly higher than the correlation coefficients obtained with Verbal IQ. There was also a trend with Buck's system for the correlation coefficients obtained with Performance IQ to be higher than the correlation coefficients obtained with Full Scale IQ.

Table 10

The t-test Values Obtained for the Difference in the Pearson Correlation Coefficients of Buck's (1948) Scoring System with Performance, Verbal and Full Scale IQ (n=101)

Dependent Correlation Coefficients	t	p
Buck's (1948) raw score with Performance IQ (r = .403) to Buck's (1948) raw score with Verbal IQ (r = .164)	2.48	< .01
Buck's (1948) raw score with Performance IQ (r = .403) to Buck's (1948) raw score with Full Scale IQ (r = .326)	1.64	< .10
Buck's (1948) weighted score with Performance IQ (r = .432) to Buck's (1948) weighted score with Verbal IQ (r = .181)	2.65	< .01
Buck's (1948) weighted score with Performance IQ (r = .432) to Buck's (1948) Weighted score with Full Scale IQ (r = .356)	1.44	< .10
Buck's (1948) standard score with Performance IQ (r = .383) to Buck's (1948) standard score with Verbal IQ (r = .150)	2.40	< .01

(continued)

Table 10 (continued)

Dependent Correlation Coefficients	t	p
Buck's (1948) standard score with Performance IQ (r = .383) to Buck's (1948) standard score with Full Scale IQ (r = .305)	1.54	< .10

Table 11

The t-test Values Obtained for the Difference in the Pearson Correlation Coefficients of the Goodenough-Harris Drawing Test with Performance, Verbal and Full Scale IQ (n=101)

Dependent Correlation Coefficients	t	p
Goodenough-Harris raw score with Performance IQ (r = .415) to Goodenough-Harris raw score with Verbal IQ (r = .101)	3.30	< .001
Goodenough-Harris raw score with Performance IQ (r = .415) to Goodenough-Harris raw score with Full Scale IQ (r = .306)	2.03	< .025
Goodenough-Harris standard score with Performance IQ (r = .369) to Goodenough-Harris standard score with Verbal IQ (r = .110)	2.66	< .01
Goodenough-Harris standard score with Performance IQ (r = .369) to Goodenough-Harris standard score with Full Scale IQ (r = .288)	1.48	< .10

For the Goodenough-Harris Drawing Test, the correlation coefficients obtained with Performance IQ were significantly higher than the correlation coefficients obtained with Verbal IQ. The correlation coefficient obtained between the Goodenough-Harris raw score and Performance IQ was also significantly higher than the correlation coefficient obtained between the Goodenough-Harris raw score and Full Scale IQ. There was a trend for the correlation coefficient of the Goodenough-Harris standard score and Performance IQ to be higher than the correlation coefficient of the Goodenough-Harris standard score and Full Scale IQ.

Hypothesis #5

The Draw-A-Person test requires that subjects take a blank piece of paper, and form a human figure with few environmental cues to guide them. For this reason, the fifth hypothesis predicted that the factor of field independence, as developed by Witkin (1965), and measured by the sum of the Picture Completion, Block Design, and Object Assembly subtests, would serve as a superior predictor of cognitive scores on the Draw-A-Person test, than would Verbal IQ, Performance IQ, or Full Scale IQ.

To test this hypothesis, a number of multiple regression analyses were performed, with the raw and standard scores of the Goodenough-Harris Drawing Test, as well as the raw, weighted, and standard scores of Buck (1948)'s scoring system, each serving as the dependent or

criterion variables. For every one of the regression analyses, the 11 subtest scores of the Wechsler Adult Intelligence Scale-Revised, and the Verbal IQ, Performance IQ, and Full Scale IQ, all served as independent or predictor variables.

The regression analyses were carried out with the SPSS-X Batch system of data analysis, and a stepwise selection of independent variables was performed. The stepwise procedure selected independent variables through a process of both forward selection and backwards elimination. With forward selection, the F test was calculated for the hypothesis that the coefficient of the entered variable was 0. An independent variable was only put into the equation if the probability of the F statistic was less than the criterion value of .05. After each step in the selection process, the variables already in the equation were then examined for possible elimination. With backwards elimination, the selected independent variables were removed unless the probability of the F value was less than .10.

Once the stepwise selection procedure had determined the multiple correlation coefficients, these coefficients were adjusted with a shrinkage formula developed by Wherry, and first reported by Lord and Novick (1968). This formula has been recommended by Carter (1979), as well as Glass and Hopkins (1984), as a simple way to eliminate the cumulative sampling error that results from the use of multiple

predictor variables. The results of the regression analyses, with both the original multiple correlation coefficients, and the adjusted correlation coefficients, are reported in Tables 12 to 16. The results did not support the predicted hypothesis, since the subtest scores that constitute field independence (Block Design, Picture Completion, and Object Assembly) did not emerge as superior predictors of cognitive scores on the Draw-A-Person test.

Table 12

Results of Regression Procedure with Goodenough-Harris Raw Score as Dependent Variable (n=101)

Step	Model	r	r^2	Adj.r	Adj.r^2	F
1	G-H raw = Picture Arrangement	.425	.181			21.83***
2	G-H raw = Picture Arrangement Object Assembly	.513	.263	.499	.249	10.97***

* p. < .05
** p. < .01
*** p. < .001

Table 13

Results of Regression Procedure with Goodenough-Harris Standard Score as Dependent Variable (n=101)

Step	Model	r	r^2	Adj.r	Adj.r^2	F
1	G-H standard = Picture Arrangement	.396	.157			18.44***
2	G-H standard = Picture Arrangement Object Assembly	.471	.222	.454	.206	8.15**

* p. < .05
** p. < .01
*** p. < .001

Table 14

Results of Regression Procedure with Buck's (1948) Raw Score as Dependent Variable (n=101)

Step	Model	r	r^2	Adj.r	Adj.r^2	F
1	Buck raw = Picture Completion	.437	.191			23.36***
2	Buck raw = Picture Completion Object Assembly	.488	.239	.473	.223	6.14*
3	Buck raw = Picture Completion Object Assembly Picture Arrangement	.522	.272	.500	.250	4.47*

* p. < .05
** p. < .01
*** p. < .001

Table 15

Results of REgression Procedure with Buck's (1948) Weighted Score as Dependent Variable (n=101)

Step	Model	r	r^2	Adj.r	Adj.r^2	F
1	Buck weighted = Performance IQ Score	.432	.187			22.73***
2	Buck weighted = Performance IQ Score Digit Symbol	.495	.229	.479	.230	7.53**

* p. < .05
** p. < .01
*** p. < .001

Table 16

Results of Regression Procedure with Buck's (1948) Standard Score as Dependent Variable (n=101)

Step	Model	r	r^2	Adj.r	Adj.r^2	F
1	Buck standard = Picture Completion	.387	.150			17.44***
2	Buck standard = Picture Completion Object Assembly	.450	.203	.423	.186	6.48*
3	Buck standard = Picture Completion Object Assembly Picture Arrangement	.490	.240	.470	.216	4.74*

* p. < .05
** p. < .01
*** p. < .001

DISCUSSION

This study attempted to answer a number of questions about intellectual evaluation based on human figure drawings by exploring the relationship of both Buck's (1948) system and the Goodenough-Harris Drawing Test to the Wechsler Adult Intelligence Scale-Revised. In order to discuss the study's results in a meaningful fashion, the inter-rater reliability of the two Draw-A-Person scoring systems will first be examined. Information concerning the validity of Buck's scoring system will then be discussed, followed by a discussion of information concerning the validity of the Goodenough-Harris Drawing Test. Once the validity of these two scoring systems has been evaluated, their potential utility as assessment tools will be compared and contrasted, in light of the study's findings. Finally, the present study's implications for future research will be discussed.

Inter-Rater Reliability

If a measure of cognitive ability is to have any usefulness, it must at the very least, have adequate inter-rater reliability. In other words, if two clinicians are shown the same sample of a subject's behavior, then these clinicians should arrive at similar estimates of the subject's cognitive skills. For this reason, it was important for the present study to examine the inter-rater

reliability of both the Goodenough-Harris Drawing Test and Buck's (1948) system, even though no specific hypotheses about reliability were tested.

The results in this area were quite encouraging. With a minimal amount of practice (a few one hour sessions), the two raters in the present study were able to achieve a high percent of inter-rater agreement for both the Goodenough-Harris Drawing Test and Buck's (1948) system. For the Goodenough-Harris drawing test, these findings are in keeping with past research. Harris (1963) reported that past inter-rater correlation coefficients were typically above .90 for the Goodenough-Harris system. While Buck (1948,1966) did not report inter-rater reliability figures for his system, the present study suggests that his method is relatively easy to learn and can yield high inter-rater reliability.

Validity of Buck's (1948) Scoring System

Given that a high percent of inter-rater agreement was obtained with Buck's (1948) system, one can begin to wonder if scores yielded by the system are an accurate reflection of intellectual ability. The present study hypothesized that scores yielded by Buck's scoring system would be significantly and positively correlated with Verbal, Performance, and Full Scale IQ scores.

The results of the study provided partial support for this hypothesis. The raw, weighted, and standard scores on

Buck (1949)'s system were significantly correlated with both Performance and Full Scale IQ. While the raw, weighted, and standard scores were not significantly correlated with Verbal IQ, there was a trend for the raw and weighted scores to be correlated with Verbal IQ.

Initially, such correlation coefficients, particularly those obtained for Buck's (1948) scoring system with Performance and Full Scale IQ, suggest a strong relationship. The astute clinician, however, would probably do well to interpret these findings with great caution. Because of the relatively large number of subjects in the study (n=101), modest correlations that only account for a small percent of the variance were statistically significant. It is also dangerous to assume that if two scores are significantly correlated, such as Buck's scores and Performance IQ, that clinicians can predict one score from the corresponding score on the other system. Buck's scoring technique is a case in point, since further analyses revealed that in this study Buck's standard scores were significantly lower than both WAIS-R Performance IQ and Full Scale IQ. The results of this study must therefore be interpreted with caution, even though they do suggest a significant relationship between Buck's scores and both Performance and Full Scale IQ.

It is also interesting to note, that while the correlation coefficients for Buck's (1948) scores were

statistically significant, they were generally lower than the Pearson correlation coefficients that Buck (1966) reported in his revised manual for the H-T-P technique. In this publication, Buck (1966) stated that in a study of 100 Caucasian adults at a state home for the mentally retarded, the following Pearson correlation coefficients for the H-T-P technique with the Wechsler-Bellevue Intelligence Scale were obtained: for the H-T-P and Verbal IQ, $\underline{r}$ = .699, for the H-T-P and Performance IQ, $\underline{r}$ = .724, and for the H-T-P and Full Scale IQ, $\underline{r}$ = .746.

Several factors could explain why these coefficients were higher than those obtained in the present study. First, it may be that Buck's (1948) system has a relatively low ceiling for the assessment of cognitive skills. This would explain why the system functioned more effectively with a group of mentally retarded adults than with a group of college students, since the retarded adults had lower cognitive skills. Buck's system may lack the ability to assess average to superior levels of intelligence with much accuracy. Or it may be that the additional data, which Buck (1966) obtained from the house and tree drawings, significantly enhanced the effectiveness of his system. While the present study sought to explore the usefulness of the human figure drawing as a single measure of cognitive ability, it may be that additional drawings can enhance the validity of the human figure drawing. Finally, it could

also be the case that Buck's (1948) system was more closely related to the Wechsler-Bellevue Scale, than it is to the more contemporary WAIS-R. These different factors are all things that can be explored by future researchers to further the existing knowledge about Buck's (1948) scoring system.

Future researchers may also want to determine if Buck's (1948) scoring system has any substantial validity in the assessment of Verbal and Full Scale IQ, or if it is primarily a measure of Performance IQ. Buck (1966) reported that his system is an effective way of estimating Verbal, Performance and Full Scale IQ. The present study, however, hypothesized that Buck's scoring system would be significantly more correlated with Performance than with either Verbal or Full Scale IQ. The results provided tentative support for the presents study's hypothesis. Buck's (1948) scoring system was significantly more correlated with Performance IQ scores than with Verbal IQ scores, and there was a trend for Buck's system to be more correlated with Performance IQ scores than with Full Scale scores. These findings suggest that Buck's test is more a measure of non-verbal ability than of verbal skills.

The present study's regression analysis also indicated that various components of non-verbal IQ were the best predictors of behavior on Buck (1948)'s scoring system. It was hypothesized that the factor of field independence, as measured by the sum of the Picture Completion, Block Design,

and Object Assembly subtests, would serve as a superior predictor of cognitive scores on Buck's (1948) system. This was not the case. For Buck's raw and standard scores, the subtests of Picture Completion, Object Assembly, and Picture Arrangement were the best predictor variables. These subtests have been linked to a rather disparate group of non-verbal abilities: Picture Completion has often been considered a measure of concentration and attention to details in the environment; Object Assembly has been thought to evaluate a subject's ability to assemble material drawn from life into meaningful whole; finally Picture Arrangement has been considered a test of planning and social intelligence (Matarazzo, 1972). To make matters even more complicated, Performance IQ scores and Digit Symbol, which has been linked to short-term memory and attention (Matarazzo, 1972), served as the best predictors variables for Buck's weighted scores. Given the scattered nature of these results, it remains unclear what components of non-verbal intelligence are evaluated by Buck's scoring system. What the regression analysis does suggest is that various aspects of non-verbal intelligence, rather than verbal intelligence, are the best predictors of cognitive scores on Buck's system. In several ways, the present study has suggested that Buck's system is primarily a measure of Performance IQ. Since this finding is contrary to past

research, further investigation with other populations is needed.

Validity of the Goodenough-Harris Drawing Test

As this study addressed the meaning of Buck's (1948) scoring system, an attempt was also made to explore the potential validity of the Goodenough-Harris Drawing Test with adult subjects. It was hypothesized that the Goodenough-Harris Drawing Test would be significantly and positively correlated with Verbal IQ, Performance IQ, and Full Scale IQ.

The results of the study provided only partial support for this hypothesis. The Goodenough-Harris scores were significantly correlated with Performance IQ and Full Scale IQ, but not with Verbal IQ. As with Buck's (1948) system, the significant correlations between the Goodenough-Harris scores and WAIS-R scores should be interpreted with caution. Initially, the findings suggest a strong relationship between Goodenough-Harris scores and both Performance and Full Scale IQ. Unfortunately, with the study's relatively large sample size (n=101), modest correlation coefficients that only account for a small percent of the variance were statistically significant. Further analyses revealed as well that the Goodenough-Harris standard scores were significantly lower than either the Performance or the Full Scale IQ scores. This suggests that clinicians attempting to estimate adult IQ with the Goodenough-Harris system, will

be likely to underestimate the ability of their subjects. The results of the study must therefore be interpreted with caution, even though they do indicate a significant relationship between Goodenough-Harris scores and both Performance IQ and Full Scale IQ.

In the past, researchers who explored the relationship between the Goodenough-Harris system and the Wechsler Intelligence Scale for Children reported correlation coefficients that ranged from .38 to .77, with the majority of these coefficients being statistically significant at or beyond the .05 level (Harris, 1963). The results of the present study tended to fall on the low end of this range of coefficients. This could be due to the fact that the Goodenough-Harris system was developed for the assessment of children, and may have a relatively low ceiling. Like Buck's (1948) system, the Goodenough-Harris Drawing Test may be a more successful measure of intellectual ability when it is used with subjects who have lower cognitive skills (such as young children and the mentally retarded). Unfortunately, it would be difficult to test this idea directly by comparing child and adult subjects, since these two groups would have to be tested with different individual IQ tests. A simpler way for future researchers to test for a possible ceiling effect would be to compare the usefulness of the Goodenough-Harris Drawing Test with mentally retarded and non-retarded adults.

Future researchers may also want to continue to examine whether or not the Goodenough-Harris system is an effective way of estimating Verbal IQ and Full Scale IQ. Past research indicated that the Goodenough-Harris Drawing Test was related to Performance IQ, Verbal IQ, and Full Scale IQ at roughly the same level. The present study, however, hypothesized that human figure drawings are a largely non-verbal measure, and that the Goodenough-Harris Drawing Test would be significantly more correlated with Performance IQ, than with either Verbal IQ or Full Scale IQ. For the most part, the results supported this hypothesis, suggesting that the Goodenough-Harris Drawing Test is related more to Performance IQ than to either Verbal IQ or Full Scale IQ.

As with Buck's (1948) system, the regression analysis also determined that several components of non-verbal IQ were the best predictors of subjects' performance on the Goodenough-Harris Drawing Test. The Picture Arrangement and Object Assembly subtests of the WAIS-R emerged as the best predictor variables for both the Goodenough-Harris raw and standard scores. As previously stated, Picture Arrangement has been considered a test of planning and social intelligence, while Object Assembly has been thought to evaluate a subject's ability to assemble material drawn from life into meaningful whole (Matarazzo, 1972). The Goodenough-Harris Drawing Test may thus be a measure of how well an individual can notice the appropriate physical

features and clothing of other people, and then turn this knowledge into a carefully planned and executed human figure drawing. In any case, it seems that non-verbal skills best predict a subject's performance on the Goodenough-Harris scoring system, or vice-versa.

Buck vs. Goodenough-Harris: A Comparison

Given the large number of features that overlap between the Goodenough-Harris Drawing Test and Buck's (1948) system, this study predicted that the two tests would be related in a similar fashion to subjects' performance on a standard, individually administered IQ test. The results supported this hypothesis. No significant difference was found between the way in which Buck's (1948) system and the Goodenough-Harris system were correlated with Full Scale IQ. It was also true that the standard scores of both systems were significantly lower than Full Scale IQ scores. While there was some difference in which subtests served as superior predictors in the regression analysis for the Goodenough-Harris Drawing Test and Buck's system, the overall relationships between the two human figure drawing systems and Full Scale IQ were remarkably similar.

This similarity was present, despite the fact that the Goodenough-Harris Drawing Test was developed for the assessment of children, while Buck's (1948) system was developed for the assessment of adults. Part of this finding may be due to the fact that Buck developed his

system at a state home for the mentally retarded. It may be that both Goodenough-Harris (1963) and Buck (1948) developed similar systems that assess low-level, undeveloped cognitive skills. The systems will thus work in a similar fashion whether they are given to adults or children. Since the present study demonstrated that the two cognitive scoring systems provide similar estimates of adult IQ, it would be useful for future researchers to test both systems with children. This would help determine if the age of a subject affects the potential validity of either the Goodenough-Harris Drawing Test or Buck's system.

Suggestions for Future Research

This study raises a number of possible questions for further research. Several of them are stated and discussed below.

First, in keeping with past research, the present study found high levels of inter-rater reliability for both the Goodenough-Harris drawing test and Buck's (1948) system. Unfortunately, the present study was not able to investigate test-retest reliability, since human figure drawings were only obtained on one occasion. While high levels of test-retest reliability were obtained for the Goodenough-Harris Drawing Test in the past, this research was conducted several decades ago and was only done with children. Contemporary researchers would do well to reexamine the test-retest reliability of both Buck's (1948) system and the

Goodenough-Harris Drawing Test with adult subjects.

In addition, past research (Buck, 1966) yielded higher correlation coefficients between the H-T-P and individual IQ tests, than were obtained in the present study. Buck (1966) may have obtained his superior results for several reasons. Buck (1966) studied the house, tree, and person drawings of mentally retarded subjects, while the present study examined only the person drawings of undergraduate volunteers. It may be the case that Buck (1948)'s H-T-P system has a relatively low ceiling and is thus more effective with mentally retarded subjects, or it may be that a combination of house, tree, and person drawings can produce a more valid estimate of IQ than just a person drawing. These ideas could be tested by conducting a research project that obtains house, tree, and person drawings from both mentally retarded and non-retarded adult subjects. This would allow researchers to compare the effectiveness of Buck (1948)'s system with mentally retarded and non-retarded adults, and also to compare the effectiveness of person drawings against the combined effectiveness of house, tree, and person drawings.

There is also some indication that the Goodenough-Harris Drawing Test may suffer from a ceiling effect. For this reason, it is recommended that future researchers compare the effectiveness of the Goodenough-Harris Drawing Test with mentally retarded and non-retarded adults.

Further, the present study found that cognitive scores on both Buck (1948)'s scoring system and the Goodenough-Harris drawing test were significantly more correlated with Performance IQ than with Verbal IQ, and that there was generally a trend for both Buck's scoring system and the Goodenough-Harris drawing test to be more correlated with Performance IQ than with Full Scale IQ. This finding was not in keeping with the results of previously published research. It is therefore recommended that the two systems of scoring human figure drawings be tested with other populations, to determine if they are in fact more related to Performance IQ scores, than to either Verbal or Full Scale IQ scores.

Finally, the present study demonstrated that the Buck's (1948) scoring system and the Goodenough-Harris Drawing Test were correlated in a similar fashion with the scores of adult subjects on an individually administered IQ test. It would thus be helpful for future researchers to test both Buck's system and the Goodenough-Harris system with children. This would help determine if the age of a subject affects the potential validity of either the Goodenough-Harris Drawing Test of Buck's system.

REFERENCES

Adler, P.T. (1971). Ethnic and socioeconomic status differences in human figure drawings. Journal of Consulting and Clinical Psychology, 36, 344-354.

Anastasi, A. (1982). Psychological Testing (fifth edition). New York: MacMillan Publishing Co.

Ayres, A.J., & Reid, W. (1966). The self-drawing as an expression of perceptual-motor dysfunction. Cortex, 2, 154-265.

Basow, S. (1986). Gender stereotypes: Traditions and alternatives. Pacific Grove, CA: Brooks Cole Publishing Company.

Bauer, W.D., & Paludi, M.A. (1979). Impact of sex of experimenter on the Draw-A-Person test. Perceptual and Motor Skills, 49, 456-458.

Beck, M., & Bart, L. (1970). Inter-rater and test-retest reliability of a proportionality measure for the D-A-P. Perceptual and Motor Skills, 30, 89-90.

Buck, J.C. (1948). The H-T-P test. Journal of Clinical Psychology, 4, 151-159.

Buck, J.N. (1966). The House-Tree-Person Technique-A revised manual. Beverly Hills, CA: Western Psychological Services.

Burt, C. (1921). Mental and scholastic tests. London: P.S. King and Sons.

Calhoun, G., Ross, J., & Bolton, J. (1988). Relationship between human figure drawings and self-esteem. Perceptual and Motor Skills, 66, 253-254.

Carter, E.S. (1979). Comparison of different shrinkage formulas in estimating population multiple correlation coefficients. Educational and Psychological Measurements, 39, 261-266.

Cassel, R.H., Johnson, A., & Burns, W.H. (1958). Examiner, ego defense, and the H-T-P Test. Journal of Clinical Psychology, 14, 157-160.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 225-251.

Cooke, E. (1885). Art teaching and child nature. London Journal of Education, 2, 10-11.

Cox, M.V., & Hawarth, C. (1989). The Human figure drawings of normal children and those with severe learning disabilities. British Journal of Developmental Psychology, 7, 333-339.

Cramer-Azima, F.J. (1956). Personality changes and figure drawings: A case treated with ACTH. Journal of Clinical Psychology, 20, 143-149.

Craddick, R. (1962). Draw-A-Person characteristics of psychopathic prisoners and college students. Perceptual

and Motor Skills, 15, 11-13.

Cressen, R. (1975). Artistic quality of drawings and judges' evaluations of the DAP. Journal of Personality Assessment, 39, 132-137.

Datta, L. & Drake, A. (1968). Examiner sex and sexual differentiation in pre-school children's figure drawings. Journal of Projective Techniques and Personality Assessment, 32, 397-399.

Daum, J.M. (1983). Emotional indicators in drawings of aggressive or withdrawn male deliquents. Journal of Personality Assessment, 47, 243-249.

Dauod, F.S. (1976). First-drawn pictures: A cross-cultural investigation. Journal of Personality Assessment, 40, 376-377.

Delatte, J.G. (1985). Significance of femininity in human figure drawings of girls. Psychological Reports, 56, 165-166.

Delatte, J.G., & Hendrickson, N.J. (1982). Human figure drawings and self-esteem. Journal of Personality Assessment, 46, 603-606.

Duffy, K.G., Beaty, J.W., & DeJulio, S. (1982). Size of human figure drawings as influenced by instructions for "sexy" versus "average" drawings and by status of the experimenter. Journal of Clinical Psychology, 38, 191-197.

Feher, E., Vandecreek, L., & Teglasi, H. (1983). The Problem of art quality in human figure drawing tests. Journal of Clinical Psychology, 39(2), 268-275.

Feldman, M., & Hunt, R.G. (1958). A relations of difficulty in drawing and ratings of adjustment based on human figure drawings. Journal of Consulting Psychology, 22, 217-220.

Ginzburg, B.M., Mersky, H., and Lau, C.L. (1988). The Relationship between pain drawings and the psychological state. Pain, 35, 141-146.

Glass, G.V., & Hopkins, K. (1984). Statistical Methods in Education and Psychology. Englewood Cliffs, N.J.: Prentice-Hall.

Goodenough, F.L. (1926). Measurement of intelligence by drawings. New York: Harcourt, Brace, and World, Inc.

Goodenough, F.L., & Harris, D.B. (1963). The Goodenough-Harris Drawing Test. New York: Harcourt, Brace, and World, Inc.

Guinan, J.F., & Hurley, J.R. (1965). An Investigation of the reliability of human figure drawings. Journal of Projective Techniques, 29, 300-304.

Hamilton, L.S. (1984). Human figure drawings as measures of of self-concept development in bilingual children. Journal of Instructional Psychology, 11, 27-36.

Hammer, E.F. (1954). Guide for qualitative research with the H-T-P. Journal of General Psychology, 51, 41-60.

Hammer, E.F. (1958). The Clinical application of projective drawings. Springfield, IL: Charles C. Thomas.

Handler, L. (1967). Anxiety indexes in the Draw-A-Person: A scoring manual. Journal of Projective Techniques and Personality Assessment, 31, 46-57.

Harris, D.B. (1963). Children's drawings as measures of intellectual maturity. New York: Harcourt, Brace, and World, Inc.

Hays, W.L. (1981). Statistics. New York: Holt, Rinehart, and Winston.

Holzberg, J.D., & Wexler, M. (1950). The validity of human form drawings as a measure of personality deviation. Journal of Projective Techniques, 14, 343-361.

Holmes, C.B., & Wiederbolt, J. (1982). Depression and figure size on the Draw-A-Person test. Perceptual and Motor Skills, 55, 825-826.

Holtzman, W.H. (1952). The examiner as a variable in the Draw-A-Person test. Journal of Consulting Psychology, 16, 1450148.

Hosteling, H. (1940). The selection of variates for use in prediction with some comments on the general problem of nuisance parameters. Annals of Mathematical Statistics, 11, 271-283.

Hozier, A. (1959). On the breakdown of the sense of reality: A study of spatial perception in schizophrenia. Journal of Clinical Psychology, 33, 185-194.

Janzen, W.B., & Coe, W.C. (1975). Clinical and sign prediction: The Draw-A-Person and female homosexuality. Journal of Clinical Psychology, 31, 757-765.

Jensen, D.E., Prandoni, J.R., & Abudabbeh, N.N. (1971). Figure drawings by sex offenders and a random sample of offenders. Perceptual and Motor Skills, 32, 295-300.

Jensen, K.W. (1985). Sex of the administrator as a variable affecting drawing a person. Perceptual and Motor Skills, 60, 72-74.

Johnson, F.A., & Greenberg, R.P. (1978). Quality of drawing as a factor in the interpretation of figure drawings. Journal of Personality Assessment, 42, 489-495.

Johnston, F.A., & Johnston, S.A. (1986) Differences between human figure drawings of child molesters and control groups. Journal of Clinical Psychology, 42, 638-647.

Jolles, I. (1964). A Catalogue for the Qualitative Interpretation of the House-Tree-Person. Beverly Hills, CA: Western Psychological Services.

Kahill, S. (1984). Human figure drawings in adults: An update of the empirical evidence. 1967-1982. Canadian Psychology/Psychologie Canadienne, 25, 269-292.

Kay, S.R. (1978). Qualitative differencesin human figure drawings according to schizophrenic subtype. Perceptual and Motor Skills, 47, 923-932.

Kirchner, J.H., & Marzolf, S.S. (1974). Personality of alcoholics as measured by Sixteen Personality Factor Questionnaire and House-Tree-Person color-choice characteristics. Psychological Reports, 35, 627-642.

Koppitz, E.M. (1968). Psychological evaluation of children's human figure drawings. New York: Grune and Stratton.

Koppitz, E.M. (1982). Personality assessment in the schools. In C.R. Reynolds & T.B. Gutkin (Eds.) The handbook of school psychology. New York: Wiley. 273-295.

Koppitz, E.M., & Casullo, M.M. (1983). Exploring cultural influences on human figure drawings of young adolescents. Perceptual and Motor Skills, 57, 479-483.

Kurdek, L.A., & Darnell-Goetschel, G. (1987). Young adolescents' human figure drawings as indicators of psychopathology. Journal of Adolescenct Research, 2, 69-74.

Levy, A.J., & Barowsky, E.I. (1986). Comparison of computer-administered Harris Goodenough Draw-A-Man test with standard paper-and-pencil administration. Perceptual and Motor Skills, 63, 395-398.

Lord, F.M., & Novick, M.R. (1968). Statistical Theories of Mental Test Scores. Boston: Addison-Wesley.

Lubin, B., Wallis, R.R., & Paine, C. (1971). Patterns of psychological test usage in the United States: 1935-1969. Journal of Professional Psychology, 2, 70-74.

Machover, K. (1949). Personality projection in the drawing of the human figure. Springfield, Illinois: Charles C. Thomas.

Maloney, M.P., & Glasser, A. (1982). An evaluation of the clinical utility of the Draw-A-Person test. Journal of Clinical Psychology, 38, 183-190.

Maloney, M.P., & Payne, L.E. (1969). Validity of the Draw-A-Person test as a measure of body image. Perceptual and Motor Skills, 29, 199-122.

Marzolf, S.S., & Kirchner, J.H. (1970). Characteristics of House-Tree-Person drawings of college men and women. Journal of Projective Techniques and Personality Assessment, 34, 138-145.

Matarazzo, J.D. (1972). Wechsler's measurement and appraisal of adult intelligence. Baltimore: Williams and Wilkins.

Melikian, L.H. (1972). First drawn picture and modernization. Journal of Personality Assessment, 36(6), 576-580.

Miller, T.M., Veltkamp, L.J., & Janson, D. (1987). Projective measures in the clinical evaluation of sexually abused children. Child Psychiatry and Human Development, 18(1), 47-57.

Montaque, D.J., & Prytula, R.E. (1975). Human figure drawing characteristics related to juvenile deliquents. Perceptual and Motor Skills, 40, 623-630.

Moses, S. (1990, November). Unusual coalition nixes inclusion of ADD in bill. APA Monitor, p.37.

Munroe, R.L., & Munroe, R.H. (1983). Drawings and values in three east african societies. The Journal of Social Psychology, 119, 135-136.

Ottenacher, K., Haley, D., Abbot, C., & Watson, P. (1984). Human figure drawing ability and vestibular processing dysfunction in learning-disabled children. Journal of Clinical Psychology, 40, 1084-1089.

Paine, P., Alves, E., & Tubino, P. (1985). Size of human figure drawings and Goodenough-Harris scores of pediatric oncology patients: a pilot study. Perceptual and Motor Skills, 60, 911-914.

Paludi, M.A. (1978). Machover revisited: Impact of sex-role orientation on sex sequence on the Draw-A-Person test. Perceptual and Motor Skills, 47, 713-714.

Pfeffer, K. (1984). Interpretation of studies of ethnic identity: Draw-A-Person as a measure of ethnic identity. Perceptual and Motor Skills, 59, 835-838.

Pfeffer, K. (1987). Effects of instructions to subjects on Draw-A-Person as a measure of ethnic identity. Perceptual and Motor Skills, 64, 780-782.

Pfeffer, K., & Olowu, A. (1986). Effects of socioeconomic differences on the sophistication of nigerian children's human figure drawings. Perceptual and Motor Skills, 62, 771-774.

Ponzo, E. (1957). An experimental variation of the Draw-A-Person technique. Journal of Projective Techniques, 21, 278-285.

Poster, E.C., Betz, C., McKenna, A., & Mossar, M. (1986). Children's attitudes toward the mentally ill as reflected in human figure drawings and stories. Journal of the American Academy of Child Psychiatry, 25, 680-686.

Prewett, P., Bardos, A.N., & Naglieri, J.A. (1988). Use of the Matrix Analogies Test-Short Form and the Draw-A-Person: A Qualitative Scoring System with learning-disabled and normal students. Journal of Psychoeducational Assessment, 6, 347-353.

Przybyla, D.P., Byrne, D., & Allgeier, E.R. (1988). Sexual attitudes as correlates of sexual details in human figure drawing. Archives of Sexual Behavior, 17, 99-105.

Pustel, G. Sternlicht, M., & Deutsch, M. (1971). Feminine tendencies in figure drawings by male homosexual retarded dyads. Journal of Clinical Psychology, 27, 260-261.

Rierdan, J., & Koff, E. (1980). Representation of the female body by early and late adolescent girls. Journal of Youth and Adolescence, 9, 339-346.

Rierdan, J., Koff, E., & Keller, H. (1982). Gender, anxiety, and human figure drawings. Journal of Personality

Assessment, 46(6), 594-596.

Ricci, C. (1894). The Art of children. (D. Maitland, Trans.) Pedagogical Seminar, 3, 302-307. (Original work published in 1887).

Richter, L.M., Griesel, R.D., and Wortley, M.E. (1989). The Draw-A-Man test: A 50-year perspective on drawings done by black South African children. South African Journal of Psychology, 19, 1-5.

Roback, H.B. (1968). Human figure drawings: Their untility in the clinical psychologist's aramentarium for personality assessment. Psychological Bulletin, 70, 1-19.

Rubin, J.A., Schacter, J., & Ragins, N. (1983). Intraindividual variability in human figure drawings: A developmental study. American Journal of Orthopsychiatry, 53, 654-667.

Saarni, C., & Azaza, V. (1977). Developmental analyses of human figure drawings in adolescence, young adulthood, and middle age. Journal of Personality Assessment, 41, 31-38.

Schaefer, C.E. (1972). Primary process elements in the Draw-A-Person protocols of creative young women. Perceptual and Motor Skills, 35, 245-246.

Scofield, J.W. (1978). An exploratory study of the Draw-A-Person as a measure of ethnic identity. Perceptual and Motor Skills, 46, 311-321.

Seifert, C.D. (1988). The Human-Figure Drawing in the treatment of an autistic adolescent. Child Psychiatry and Human Development, 19, 74-81.

Shaffer, J.W., Pearson, T.A., Mead, L.A., & Thomas, C.B. (1986). A Possible relationship between figure drawing data collected in medical school and later health status among physicians. Journal of Clinical Psychology, 42, 363-369.

Short-DeGraff, M.A., Slansky, L., & Diamond, K.E. (1989). Validity of preschoolers' self-drawings as an index of human figure drawing performance. The Occupational Therapy Journal of Research, 9, 305-315.

Sidun, N.M., & Rosenthal, R.H. (1987). Graphid indicators of sexual abuse in Draw-A-Person tests of psychiatrically hospitalized adolescents. The Arts in Psychotherapy, 14, 25-33.

Soccolich, C., & Wysocki, B.A. (1967). Draw-A-Person protocols of male and female college students. Perceptual and Motor Skills, 25, 873-879.

Solar, D., Bruehl, D., & Kovacs, J. (1970). The Draw-A -Person test: Social conformity or artistic ability? Journal of Clinical Psychology, 26, 524-525.

Strumpfer, D. J. (1963). The Relation of Draw-A-Person Test variables to age and chronicity in psychotic groups. Journal of Clinical Psychology, 19, 208-211.

Swensen, C.H. (1957). Empirical investigations of human figure drawings. Psychological Bulletin, 54, 431-466.

Swensen, C.H. (1968). Empirical evaluations of human figure drawings: 1957-1966. Psychological Bulletin, 70, 20-44.

Sundberg, N.D. (1961). The practise of psychological testing in clinical service in the United States. American Psychologist, 16, 79-83.

Taylor, P. (1977). An Investigation of the utility of the draw-a-person-the-rain for the assessment of stress and the prediction of achievement in college students (Doctoral dissertation, University of Cincinnati, 1977). Dissertation Abstracts International, 38/02A, 712.

Teglasi, H. (1980). Acceptance of the traditional female role and sex of the first person drawn on the Draw-A-Person test. Perceptual and Motor Skills, 51, 267-271

Thomas, C.B., Jones, L.W., & Ross, D.C. (1968). Studies on figure drawings: Biological implications of structural and graphic characteristics. Psychiatric Quarterly Supplement, Part 2, 1-29.

Wagoner, M.E., & Schubert, H.J. (1955). D.A.P. quality scale for late adolescents. Kenmore, New York: Deleware Letter Shop.

West, J.V., Baugh, V.S., & Baugh, A.P. (1963). Rorschach and Draw-A-Person responses of hypnotized and non-

hypnotized subjects. Psychiatric Quarterly, 37, 123-127.

Wildman, R.W., & Wildman, R.W. II. (1967). The practise of clinical psychology in the United States. Journal of Clinical Psychology, 31, 455-458.

Witkin, H.A. (1965). Psychological differentiation and forms of pathology. Journal of Abnormal Psychology, 70, 317-336.

Wysocki, A.C., & Wysocki, B.A. (1977). Human figure drawings of sex offenders. Journal of Clinical Psychology, 33, 224-226.

Yates, A., Beutler, L.E., Crago, M. (1985). Drawings by child victims of incest. Child Abuse and Neglect, 9, 183-189.

Zucker, K.J., Finegan, J.K., Doering, R.W., & Bradley, S. (1983). Human figure drawings of gender-problem children: A comparison to sibling, psychiatric, and normal controls. Journal of Abnormal Child Psychology, 11, 287-298.

APPROVAL SHEET

The dissertation submitted by Steven Abell has been read and approved by the following committee:

Dr. James E. Johnson, Director
Professor
Loyola University Chicago

Dr. Alan S. DeWolfe
Professor
Loyola University Chicago

Dr. Richard Maier
Associate Professor
Loyola University Chicago

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval by the Committee with reference to content and form.

The dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

7-29-91
Date


Director's Signature